

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:47:52 ; Search time 40 Seconds  
(without alignments)  
2877.287 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTKMRRVGVGVGTIGG.....IIWRATEGIPKSEILRTIF 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL 23:\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phage.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*
  - 15: sp\_rvirus.\*
  - 16: sp\_bacteriap.\*
  - 17: sp\_archheap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2293	100.0	446	10	Q9LDI3	Q9ldi3 arabidopsis
2	1884.5	82.2	400	10	Q9LKR2	Q9lkr2 arabidopsis
3	1484.5	64.7	445	10	Q9STV4	Q9stv4 arabidopsis
4	1381.5	60.2	480	10	Q9C753	Q9c753 arabidopsis
5	1376.5	60.0	482	10	Q93VD3	Q93vd3 arabidopsis
6	1333	58.1	440	10	O24342	O24342 sorghum bic
7	1314	57.3	441	10	Q8RWU8	Q8rwu8 arabidopsis
8	1313	57.3	440	10	O24343	O24343 sorghum bic
9	1262.5	55.1	373	10	Q8GSY8	Q8gsy8 oryza sativ
10	1251.5	54.6	447	10	Q94F03	Q94f03 arabidopsis
11	1246.5	54.4	449	10	Q9C5S4	Q9c5s4 arabidopsis
12	1243.5	54.2	451	10	Q9C5P5	Q9c5p5 arabidopsis
13	1241.5	54.1	542	10	Q8SV9	Q8sv9 oryza sativ
14	1234.5	53.8	453	10	Q9MAM1	Q9mam1 arabidopsis
15	1202.5	52.4	445	10	Q8LIG4	Q8lig4 oryza sativ
16	1149	50.1	382	10	Q9ZVD9	Q9zvd9 arabidopsis

17	1145	49.9	375	10	Q9LKC9	Q9lkc9 arabidopsis
18	1090	47.5	461	10	Q9LWM4	Q9lwm4 oryza sativ
19	1043	45.5	438	10	Q8LX24	Q8lkx24 glycine max
20	1040	45.4	441	10	O65554	O65554 arabidopsis
21	1039.5	45.3	439	10	Q9FJ54	Q9fj54 arabidopsis
22	1032.5	45.0	441	10	Q8W2D7	Q8w2d7 brassica na
23	1032.5	45.0	502	10	Q8LRC0	Q8lrc0 oryza sativ
24	1031.5	45.0	461	10	Q9LGV5	Q9lgv5 oryza sativ
25	1021	44.5	487	10	Q94A54	Q94a54 arabidopsis
26	1018.5	44.4	488	10	Q8W1D5	Q8w1d5 arabidopsis
27	1014.5	44.2	432	10	Q94C40	Q94c40 arabidopsis
28	1012	44.1	453	10	Q8H2C2	Q8h2c2 persea amer
29	1009.5	44.0	435	10	Q8H0X3	Q8h0x3 arabidopsis
30	1009.5	44.0	445	10	Q9LEU7	Q9leu7 arabidopsis
31	1006.5	43.9	456	10	Q9LYQ8	Q9lyq8 arabidopsis
32	1005	43.8	444	10	Q8RWC9	Q8rwc9 arabidopsis
33	1003	43.7	444	10	Q9FUK2	Q9fuk2 arabidopsis
34	1001	43.7	462	10	Q9XFJ3	Q9xfj3 mesembryant
35	999.5	43.6	421	10	Q9SX61	Q9sx61 arabidopsis
36	998.5	43.5	456	10	Q9LKD0	Q9lkd0 arabidopsis
37	996	43.4	480	10	Q9LUP6	Q9lup6 arabidopsis
38	986	43.0	464	10	Q9LVL3	Q9lvl3 arabidopsis
39	985.5	43.0	508	10	Q9SLZ6	Q9slz6 oryza sativ
40	984.5	42.9	454	10	Q9LNH7	Q9lnh7 arabidopsis
41	984	42.9	483	10	Q9FJ55	Q9fj55 arabidopsis
42	983.5	42.9	489	10	Q9SN43	Q9sn43 arabidopsis
43	983	42.9	518	10	Q9SEG2	Q9seg2 zea mays (m
44	982.5	42.8	479	10	Q9C562	Q9c562 arabidopsis
45	980	42.7	474	10	Q94DT7	Q94dt7 oryza sativ

ALIGNMENTS

RESULT 1

Q9LDI3 ID Q9LDI3 PRELIMINARY; PRT; 446 AA.  
AC Q9LDI3;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Serine/threonine protein kinase SOS2 (CEL-interacting protein kinase 24).  
DE GN SOS2 OR AT5G35410.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";  
RL Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RX MEDLINE=20202704; PubMed=10725382;  
RA Liu J., Ishitani M., Halfter U., Kim C.S., Zhu J.K.;  
RT "The Arabidopsis thaliana SOS2 gene encodes a protein kinase that is required for salt tolerance.";  
RL Proc. Natl. Acad. Sci. U.S.A. 97:3730-3734(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Weinl S., Albrecht V., Kudla J.;  
RT "Molecular characterization of the CIPK gene family from Arabidopsis thaliana.";  
RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AB025611; BAA98146.1; -;  
DR EMBL; AF237670; AAF62923.1; -;  
DR EMBL; AF395081; AAK72257.1; -;  
DR EMBL; AY099621; AAM20472.1; -;  
DR EMBL; BT002138; AAN72149.1; -;  
DR HSSP; Q63450; 1A06.  
DR InterPro; IPR004041; NAF dom.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF03822; NAF; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 446 AA; 50634 MW; 99FDF3A778E1093D CRC64;

Query Match 100.0%; Score 2293; DB 10; Length 446;  
Best Local Similarity 100.0%; Pred. No. 1.7e-183;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTKMRRVKGVEVGRITGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60  
DB 1 MTKMRRVKGVEVGRITGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60  
QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD 120  
DB 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVD 120  
QY 121 AVAHCCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRITCGTPNYVAPEV 180  
DB 121 AVAHCCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRITCGTPNYVAPEV 180  
QY 181 LSGQYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240  
DB 181 LSGQYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240  
QY 241 IHRILDPNPKTRIQIQIKKDPWFLNLYVPIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
DB 241 IHRILDPNPKTRIQIQIKKDPWFLNLYVPIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
QY 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
DB 301 ERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
QY 361 MGFKSHTRNFKTRLEGLSSIKAGQLAWVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420  
DB 361 MGFKSHTRNFKTRLEGLSSIKAGQLAWVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420  
QY 421 CSKLENIWRATEGIPKSEILRTITF 446  
DB 421 CSKLENIWRATEGIPKSEILRTITF 446

ID Q9LKR2 PRELIMINARY; PRT; 400 AA.  
AC Q9LKR2;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE F6I13.1 protein (Fragment).  
GN F6I13.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA WashU;  
RT "The A. thaliana Genome Sequencing Project.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Wilson R.;  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AF262044; AAF67384.1; -;  
DR HSSP; Q63450; 1A06.  
DR InterPro; IPR004041; NAF dom.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF03822; NAF; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
FT NON TER 400  
SQ SEQUENCE 400 AA; 45623 MW; 6908A8A695E12BAF CRC64;  
Query Match 82.2%; Score 1884.5; DB 10; Length 400;  
Best Local Similarity 93.2%; Pred. No. 2.5e-149;  
Matches 371; Conservative 0; Mismatches 0; Indels 27; Gaps 2;  
QY 1 MTKMRRVKGVEVGRITGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60  
DB 1 MTKMRRVKGVEVGRITGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREI 60  
QY 61 SIMKIVRHPNIVRLY-----EVLASPSKIYIVLEFVTGGELFDRI----- 100  
DB 61 SIMKIVRHPNIVRLYEVCLFVSMHLRNFISEVLASPSKIYIVLEFVTGGELFDRIVRNFH 120  
QY 101 -----VHKGRLEESERKVFQQLVDVAHCHCKGVYHRDLKPNLLDNTGNLKVSDF 153  
DB 121 TCRQRPVHKGRLEESERKVFQQLVDVAHCHCKGVYHRDLKPNLLDNTGNLKVSDF 180  
QY 154 GLSALPQEGVELLRITCGTPNYVAPEVLSGQYDGSAAADIWSCGVILFVILAGYLPFSET 213  
DB 181 GLSALPQEGVELLRITCGTPNYVAPEVLSGQYDGSAAADIWSCGVILFVILAGYLPFSET 240  
QY 214 DLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQIQIKKDPWFLNLYVPIRA 273  
DB 241 DLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQIQIKKDPWFLNLYVPIRA 300  
QY 274 REEEVNLDLDIRAVFDGIEGSYVAENVVERNDEGLMNAFEMITLSQGLNLSALFDRRQD 333  
DB 301 REEEVNLDLDIRAVFDGIEGSYVAENVVERNDEGLMNAFEMITLSQGLNLSALFDRRQD 360  
QY 334 FVKRQTRFVSRREPSEIIANIEAVANSMGFKSHTRNEK 371  
DB 361 FVKRQTRFVSRREPSEIIANIEAVANSMGFKSHTRNEK 398











Db 69 IKHPNVQLYEVMASTKIFILILEYVVTGGELFDKI VNDGRMKEDARRYFQQLIHAVDYC 128

QY 126 HCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPEVLSG 183

Db 129 HSRGVYHRDLKPENLLDSYGNLKISDFGLSALSQQVRDDGLLHTSCGTPNYVAPEVLND 188

QY 184 QGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHR 243

Db 189 RGYDGATADMWSCGVVLYVLLAGYLPFDDSNLMNLKYKISSGEFNCPPWLSLGAMKLITR 248

QY 244 ILDPNPKTRIQIQGIKDPWFRLNYPVIRAREEEEEVNLLDIRAVFDGIEGSYVAENVERN 303

Db 249 ILDPNPMTRVTPQEVFEDEWFKKDYKPPVFEERDDSNMDDIDAVFKDSEEHV--TEKR 305

QY 304 DEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGF 363

Db 306 EEQPAAINAFEIISMRSRGLNLENLFDPEQEF-KRETRITLRGGANEIIEKIEEAAKPLGF 364

QY 364 KSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVVDVRKAAAGETLEYHKFKKLCISK 423

Db 365 DVQKKNYKMLNVKAGRKGNLNVATEIFQVAPSLHMVQVSKSGDITLEFHKFKKLSNS 424

QY 424 LENIIWRATE 433

Db 425 LEQVVTNNE 434

RESULT 8

O24343 PRELIMINARY; PRT; 440 AA.

AC O24343

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Serine/threonine kinase.

GN SNFL2.

OS Sorghum bicolor (Sorghum) (Sorghum vulgare).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

OX NCBI\_TaxID=4558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. TX 430;

RX MEDLINE=98145442; PubMed=9484448;

RA Annen F., Stockhaus J.;

RT "Characterization of a Sorghum bicolor gene family encoding putative protein kinases with a high similarity to the yeast SNF1 protein kinase.";

RL Plant Mol. Biol. 36:529-539(1998).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; Y12465; CAA73068.1; -.

DR HSSP; Q63450; 1A06.

DR InterPro; IPR004041; NAF\_dom.

DR InterPro; IPR000719; Prot\_kinase.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR Pfam; PF03822; NAF; 1.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Prot\_kinase; 1.

DR SMART; SM00220; S\_TKC; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 440 AA; 50476 MW; 86345E0585D975C6 CRC64;

Query Match 57.3%; Score 1313; DB 10; Length 440;

Best Local Similarity 58.3%; Pred. No. 2.1e-101;

Matches 253; Conservative 76; Mismatches 99; Indels 6; Gaps 3;

QY 2 TKKMRRVGKYEVGRTIGEGTFAKVKFARNTTDGDNVAIKIMAKSTILKNRMVDQIKREIS 61

Db 4 TKVKRRVGKYEGLRTIGEGTFAKVRFAKNTETGEPVAIKILDKEKVLRHKMQEIKREIS 63

QY 62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESRKYFQQLVDA 121

Db 64 TMKLIKHPNVVRIYEVMSGTKIYIVLEVYVTGGELFDTIANHGRMRDEARRYFQQLINA 123

QY 122 VAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPE 179

Db 124 VDYCHSRGVYHRDLKPENLLDSYGNLKVSDFGLSALSQIQIKDDGLLHTTCGTPNYVAPE 183

QY 180 VLSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKF 239

Db 184 VLEDQGYDGAMADLWSCGVILFVLLAGYLPFEDSNLMTLYKKISNAEYTFPPWTSFPAKR 243

QY 240 LIHRILDPNPKTRIQIQGIKDPWFRLNYPVIRAREEEEEVNLLDIRAVFDGIEGSYVAEN 299

Db 244 LLTRFLDPNPMTRITPEILEDEWFKKYKRPEFDEKYDTPLDVDVDAVFNDSSEHHV-- 300

QY 300 VERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVAN 359

Db 301 TEKKEEPVVLNAPELISRSAGNLGNLFDSEQEF-KRETRFTSKCPPKEIVRKIEEAAK 359

QY 360 SMGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVVDVRKAAAGETLEYHKFKYK 419

Db 360 PLGFGVQKKNYKLRLEKVKAGRKGNLNVATEILQVAPSLHMVEVRKAKGDTLEFQKFYKN 419

QY 420 LCKSLENIIWRATE 433

Db 420 LSKTLXDVVWKSED 433

RESULT 9

Q8GSY8

ID Q8GSY8 PRELIMINARY; PRT; 373 AA.

AC Q8GSY8

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Putative serine/threonine-specific protein kinase.

GN OSUNBA0086A10.12.

OS Oryza sativa (japonica cultivar-group).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.

OX NCBI\_TaxID=39947;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Nipponbare;

RX MEDLINE=22337376; PubMed=12447438;

RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J., Niimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H., Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M., Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C., Hijishita S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M., Ikeno M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K., Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I., Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y., Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M., Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai, K, Shibata M., Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K., Waki K., Yamagata H., Yamane H., Yoshiaki S., Yoshihara R., Yukawa K., Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y., Yano M., Jiang J., Gojobori T.;

RT "The genome sequence and structure of rice chromosome 1.";

RL Nature 420:312-316(2002).

DR EMBL; AP003449; BAC19913.1; -.

KW Kinase.

SQ SEQUENCE 373 AA; 42230 MW; 7873F9BE761ADB40 CRC64;

Query Match 55.1%; Score 1262.5; DB 10; Length 373;

Best Local Similarity 64.9%; Pred. No. 2.8e-97;

Matches 239; Conservative 57; Mismatches 65; Indels 7; Gaps 2;

QY 5 MRRVGKYEVRTIGEGTFAKVKFARNTTDGDNVAIKIMAKSTILKNRMVDQIKREISMK 64

Db :||||| 7 LRVGKYEVGRTIGEGTFAKVKAQNTESGESVAMKVVDRSSILKHKMADQIKREISIMK 66

QY 65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVDAVAH 124

Db :||||| 67 LVRHPNV-----VLASRKKIFILEFITGGELFDKIIRHGRLENEADARRYFQQLIDGVDF 121

QY 125 CHCKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYVAPEVLSGQ 184

Db :||||| 122 CHSKGVYHRDLKPENLLDLSQGNLKISDFGLSAWPAQGGALLRTTCGTPNYVAPEVLSHK 181

QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244

Db :||||| 182 GYDGLADATWSCGVILYVLLAGYLPFDEVDLTTLYGKIESAEYSFPFPAWFPNGAKSLIHRI 241

QY 245 LDPNPKTRIQIQIGIKKDPWFRNLNVPIRAREEEVNLDLDIRAVFDGIEGSYVAENVERND 304

Db :||||| 242 LDPNPKTRIRIEIRNDEWFKKNYEPTREIESEEVNLDVNAAFDDPEED--ADHTLDDE 299

QY 305 EGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGF 364

Db :||||| 300 AGPLTLNAFDLIILSQGLNLAALFDRRQDYDKLQNRFLSRKPAKVIIMSSMEVVAQSMGYK 359

QY 365 SHTRNEKT 372

Db :||||| 360 THIRNYKT 367

RESULT 10

Q94F03

ID Q94F03 PRELIMINARY; PRT; 447 AA.

AC Q94F03;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Similar to wpk4 protein kinase.

GN T25K16.13 OR Atr1G01140.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Lam B., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,

RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,

RA Ecker J., Theologis A., Davis R.W.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,

RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,

RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,

RA Ecker J., Theologis A., Davis R.W.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AF386999; AAK62444.1; -.

DR EMBL; AY093242; AAM13241.1; -.

DR HSSP; P24941; 1BUH.

DR InterPro; IPR004041; NAF\_dom.

DR InterPro; IPR000719; Prot\_kinase.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR Pfam; PF03822; NAF; 1.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Prot\_kinase; 1.

DR SMART; SM00220; S\_TKC; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.

KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.

SQ SEQUENCE 447 AA; 50476 MW; D6A6BF7B813760C5 CRC64;

Query Match 54.6%; Score 1251.5; DB 10; Length 447;

Best Local Similarity 54.7%; Pred. No. 3.le-96;

Matches 233; Conservative 85; Mismatches 101; Indels 7; Gaps 3;

QY 7 RVGKYEVGRTIGEGTFAKVKAQNTESGESVAMKVVDRSSILKHKMADQIKREISIMKIV 66

Db :||||| 15 RVGNYEMGRTLGECSFAKAKYAKNTVTGDQAAIKILDREKVFRHKMVEQLKREISTMKLI 74

QY 67 RHENIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVDAVAHCH 126

Db :||||| 75 KHENVVEIIEVMASKTKIYIVLELVNGGELFDKIAQQGRLEKDEARRYFQQLINAVDYCH 134

QY 127 CKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVE--LLR TTCGTPNYVAPEVLSGQ 184

Db :||||| 135 SRGVYHRDLKPENLILDANGVLKVSDPGLSARFQVRDGLLHTACGTPNYVAPEVLSDK 194

QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244

Db :||||| 195 GYDGAADVWSCGVILFVLMAGYLPFDEPNLMTLYKRIKCAEFSCPPWFSGAKRVIKRI 254

QY 245 LDPNPKTRIQIQIGIKKDPWFRNLNVPIR-AREEEVNLDLDIRAVFDGIEGSYVAENVERN 303

Db :||||| 255 LEPNPITRISIAELLEDEWFKGKYPSPFDQDDEDITDDVDAFNSKECLVTEKKEK- 313

QY 304 DEGLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGF 363

Db :||||| 314 ---PVSMNAFELISSSEFSLENLFKQAOQLVKETRTFSQRSASEIMSKMEETAKPLGF 370

QY 364 KSHTRNFKTRLEGLSSIKAKQOLAVVIEIYEVAAPSLEFMDVRKAAGETLEYHKFYKKLCSK 423

Db :||||| 371 NVKDNKYIKMKGDKSGRKQLSVATEVEFVAPSLHVVELRKTTGGDTLLEFHKFYKNFSSG 430

QY 424 LENLIW 429

Db :||||| 431 LKDVVW 436

RESULT 11

Q9C5S4

ID Q9C5S4 PRELIMINARY; PRT; 449 AA.

AC Q9C5S4;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE CBL-interacting protein kinase 9.

GN CIPK9.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21153204; PubMed=11230129;

RA Albrecht V., Ritz O., Linder S., Harter K., Kudla J.;

RT "The NAF domain defines a novel protein-protein interaction module conserved in Ca(2+)-regulated kinases.";

RL EMBO J. 20:1051-1063(2001).

CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; AF295664; AAK16684.1; -.

DR HSSP; Q63450; 1A06.

DR InterPro; IPR004041; NAF\_dom.

DR InterPro; IPR000719; Prot\_kinase.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR Pfam; PF03822; NAF; 1.

DR Pfam; PF00069; pkinase; 1.

DR ProDom; PD000001; Prot\_kinase; 1.

DR SMART; SM00220; S\_TKC; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.



DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 449 AA; 50760 MW; BF3A98AB67AB2AFB CRC64;  
Query Match 54.4%; Score 1246.5; DB 10; Length 449;  
Best Local Similarity 54.4%; Pred. No. 8.1e-96;  
Matches 233; Conservative 86; Mismatches 100; Indels 9; Gaps 4;  
QY 7 RVGKYEVRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIV 66  
Db 15 RVGNVEMGRTLGEGSFAKVKYAKNTVTGDQAAIKILDREKVFHRHKMVEQLKREISTMKLI 74  
QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCH 126  
Db 75 KHPNVVEIIEVMASKTKIYIVLELVNGGELFDKIAQQGRLKEDEARRYFQQLINAVDYCH 134  
QY 127 CKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPEVLSGQ 184  
Db 135 SRGVYHRDLKPENLILDANGVLKVSDFGLSAFSRQVREDGILLHTACGTPNYVAPEVLSDK 194  
QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKIN--AAEFSCPPWFSAEVKFLIH 242  
Db 195 GYDGAADVWSCGVILFVLMAGYLPFDEPNLMTLYKRVIRICKAEFSCPPWFSGAKRVIK 254  
QY 243 RIIDPNPKTRIQIQIKKDPWFRNLNVPIR-AREEEVNLDIDIRAVFDGIEGYSYVAENV 301  
Db 255 RIIEPNPITRISIAELLEDEWFKKGYPSPFDQDDDEDITDDVDAAFSNSKECLVTEKKE 314  
QY 302 RNDEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVNSM 361  
Db 315 K----PVSMNAFELISSSEFSLENLFKEQAQLVKKEIRFTSQRSASEIMSKMEETAKPL 370  
QY 362 GFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKLC 421  
Db 371 GFNVKDNVYKIKMGDKSGRKQLSVATEVFEVAPSLHVVELRKTGGDTLEFHKFYKNFS 430  
QY 422 SKLENIW 429  
Db 431 SGLKDVVW 438  
RESULT 12  
Q9C5P5 PRELIMINARY; PRT; 451 AA.  
AC Q9C5P5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE SOS2-like protein kinase PKS6.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Guo Y., Halfter U., Ishitani M., Zhu J.-K.;  
RT "Binding Domain and Specificity of Interaction between SOS3 Family of  
RT Calcium Sensors and SOS2 Family of Protein Kinases.";  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AF339147; AA026845.1; -.  
DR HSSP; Q63450; 1A06.  
DR InterPro; IPR004041; NAF dom.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF03822; NAF; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 451 AA; 50909 MW; A72909C7F55581A3 CRC64;  
Query Match 54.2%; Score 1243.5; DB 10; Length 451;  
Best Local Similarity 54.4%; Pred. No. 1.5e-95;  
Matches 234; Conservative 85; Mismatches 100; Indels 11; Gaps 4;  
QY 7 RVGKYEVRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIV 66  
Db 15 RVGNVEMGRTLGEGSFAKVKYAKNTVTGDQAAIKILDREKVFHRHKMVEQLKREISTMKLI 74  
QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCH 126  
Db 75 KHPNVVEIIEVMASKTKIYIVLELVNGGELFDKIAQQGRLKEDEARRYFQQLINAVDYCH 134  
QY 127 CKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPEVLSGQ 184  
Db 135 SRGVYHRDLKPENLILDANGVLKVSDFGLSAFSRQVREDGILLHTACGTPNYVAPEVLSDK 194  
QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244  
Db 195 GYDGAADVWSCGVILFVLMAGYLPFDEPNLMTLYKRICKAEFSCPPWFSGAKRVIKRI 254  
QY 245 LDNPKNKTRIQIQIKKDPWFRNLNVPIR-AREEEVNLDIDIRAVFDGIEGYSYVAENV 303  
Db 255 LEPNPITRISIAELLEDEWFKKGYPSPFDQDDDEDITDDVDAAFSNSKECLVTEKKEK- 313  
QY 304 DEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVNSMGF 363  
Db 314 --PVSMNAFELISSSEFSLENLFKEQAQLVKKETRFTSQRSASEIMSKMEETAKPLGF 370  
QY 364 KSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVRKAAGETLEYHK----FYKK 419  
Db 371 NVKDNVYKIKMGDKSGRKQLSVATEVFEVAPSLHVVELRKTGGDTLEFHKVCDSPYKN 430  
QY 420 LCSKLENIW 429  
Db 431 FSSGLKDVVW 440  
RESULT 13  
Q8S5V9 PRELIMINARY; PRT; 542 AA.  
AC Q8S5V9;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative serine/threonine kinase.  
GN OJ1015F07.8.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,  
RA Saski C., Henry D., Oates R., Simmons J.;  
RT "Rice Genomic Sequence.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; AC104427; AAM19110.1; -.  
DR HSSP; P24941; 1BUH.  
DR Gramene; Q8S5V9; -.  
DR InterPro; IPR004041; NAF dom.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF03822; NAF; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.



DR	InterPro; IPR001245; Tyr_kinase.	
DR	Pfam; PF03822; NAF; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	ProDom; PD000001; Prot_kinase; 1.	
DR	SMART; SM00220; S_TKc; 1.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.	
KW	ATP-binding; Kinase; Transferase.	
SQ	SEQUENCE 445 AA; 50952 MW; 45D0CB10F2D46FAF CRC64;	
Query Match 52.4%; Score 1202.5; DB 10; Length 445;		
Best Local Similarity 53.3%; Pred. No. 3.9e-92;		
Matches 229; Conservative 88; Mismatches 106; Indels 7; Gaps 4;		
QY	3 KMRER-VGKYEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDDQIKREIS	61
Db	10 QKVRCLGKYELGRAIGQGTFAKVRFAKNMETGDHVAIKILDKAKVKQKRLVEQIRREIC	69
QY	62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESSESRKYFQQLVDA	121
Db	70 TMKLIQHPNVVHLHEVMGSKTRIFIVLEYVMGGELHDIATSGRLKEDEARKYFQQLINA	129
QY	122 VAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQE--GVELLRTTCGTPNYVAPE	179
Db	130 VDYCHSRGVYHRDLKLENLLDGTAGNIKVSDFGLSALSEQVKADGLLHTTCGTPNYVAPE	189
QY	180 VLSGGQYDGSAAIDWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKF	239
Db	190 VIEDKGYDGAADLWSCGVILFVLLAGYLPFEDENIVSLYNKISGAQFTCPSWFSAEAKR	249
QY	240 LIHRILDPNPKTRIQIGIKKDPWFRNLNYPPIRAREEEEEVNLLDDIRAVFDGIEGSYVAEN	299
Db	250 LIARILDPNPATRIITTSQVLQDQWFKKGYESPVFDDKYYPYFHDVYDAFGDSEEKHVKEA	309
QY	300 VERNDEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEHAVAN	359
Db	310 ME---EQPTLMNAFELISLNKGLNLDNFFESDKKY-KRETRFTSQCPPKEIINRIEEAAN	365
QY	360 SMGPKSHTRNFKTRLEGLSSIKAGQIAVVEIYEIVAPSLFMVDVRKAAGETLEYHKPYKK	419
Db	366 LLGFNIQKRNYRMRMENIKEGRKGHLNIAATEVFQVAPSLHVVELKKAKGDTLEFQKFYQT	425
QY	420 LCSKLENIIW	429
Db	426 LSTQLKDVVW	435



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:47:01 ; Search time 15 Seconds  
(without alignments)  
1398.261 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTKKMRVGKYEVRTTGG.....IIWRATEGIPKSEILRTTF 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues 127863  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	701.5	30.6	576	1 SNF1_SCHPO	O74536 schizosacch
2	700.5	30.5	512	1 KI10_ARATH	Q38997 arabidopsis
3	673	29.4	611	1 SNF1_CANGA	Q00372 candida gla
4	670	29.2	552	1 AAK2_HUMAN	P54646 homo sapien
5	665	29.0	619	1 SNF1_CANTR	O94168 candida tro
6	663	28.9	633	1 SNF1_YEAST	P06782 saccharomyc
7	662	28.9	552	1 AAK2_RAT	Q09137 rattus norv
8	661.5	28.8	620	1 SNF1_CANAL	P52497 candida alb
9	655.5	28.6	550	1 AAK1_HUMAN	Q13131 homo sapien
10	653	28.5	548	1 AAK1_RAT	P54645 rattus norv
11	610	26.6	502	1 RKI1_SECC	Q02723 secale cere
12	598	26.1	1142	1 GIN4_YEAST	Q12263 saccharomyc
13	592.5	25.8	622	1 YNA3_CABEL	P45894 caenorhabdi
14	589.5	25.7	1518	1 KKK1_YEAST	P34244 saccharomyc
15	586	25.6	752	1 MRK4_HUMAN	Q96134 homo sapien
16	576.5	25.1	776	1 SN1L_RAT	Q9rlu5 rattus norv
17	566.5	24.7	779	1 SN1L_MOUSE	Q60670 mus musculu
18	563.5	24.6	776	1 MRK3_HUMAN	P27448 homo sapien
19	561.5	24.5	774	1 MRK2_MOUSE	Q05512 mus musculu
20	561.5	24.5	786	1 SN1L_HUMAN	P57059 homo sapien
21	555.5	24.2	915	1 KCC4_YEAST	P25389 saccharomyc
22	541	23.6	800	1 KIN4_YEAST	Q01919 saccharomyc
23	534.5	23.3	775	1 CDR2_SCHPO	P87050 schizosacch
24	530.5	23.1	661	1 Y537_HUMAN	O60285 homo sapien
25	520.5	22.7	408	1 ST6L_XENLA	Q91819 xenopus lae
26	517.5	22.6	407	1 STK6_XENLA	Q91820 xenopus lae
27	513	22.4	714	1 HUNK_HUMAN	P57058 homo sapien
28	510.5	22.3	353	1 ASK2_ARATH	P43292 arabidopsis
29	510	22.2	332	1 AAIP_WHEAT	Q02066 triticum ae
30	506	22.1	476	1 CHK1_HUMAN	O14757 homo sapien
31	504.5	22.0	714	1 HUNK_MOUSE	O88866 mus musculu
32	495	21.6	593	1 CDRI_SCHPO	P07334 schizosacch
33	492	21.5	476	1 CHK1_MOUSE	O35280 mus musculu

34	490.5	21.4	363	1 ASK1_ARATH	P43291 arabidopsis
35	487.5	21.3	295	1 KMLC_DICDI	P25323 dictyosteli
36	487.5	21.3	496	1 CHK1_SCHPO	P34208 schizosacch
37	487	21.2	533	1 KCCD_RAT	P15791 rattus norv
38	485	21.2	268	1 ST2C_MOUSE	Q9d2e1 mus musculu
39	485	21.2	542	1 KCCB_MOUSE	P28652 mus musculu
40	485	21.2	542	1 KCCB_RAT	P08413 rattus norv
41	484	21.1	499	1 KCCD_HUMAN	Q13557 homo sapien
42	481.5	21.0	891	1 KIN1_SCHPO	P22987 schizosacch
43	480	20.9	478	1 KCCA_HUMAN	Q9uqm7 homo sapien
44	477.5	20.8	460	1 CDS1_SCHPO	Q09170 schizosacch
45	476	20.8	508	1 CDPK_SOYBN	P28583 glycine max

ALIGNMENTS

RESULT 1  
SNF1\_SCHPO  
ID SNF1\_SCHPO STANDARD; PRT; 576 AA.  
AC O74536;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE SNF1-like protein kinase (EC 2.7.1.-).  
GN SPCC74.03C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Nurse P.;  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe.";  
RL Nature 415:871-880(2002).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SNF1 SUBFAMILY.  
CC -!- SIMILARITY: Contains 1 UBA domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC







QY 247 PNPRTIQIQQIKKDPWRLN-----YVPIRAREEBEVN-----LDDIRAV----- 287  
DB 273 VNPLNRSIHEIMQDEWFKVLDLAAYLVPQDLKQEQFNKKSNGNEENVEEIDDEMVTLSK 332  
QY 288 FDGIEGYSVAENVERNDGFL-----MMNAFEMITLSQGLNLSALFDRRDQ--FVKRQTRFV 342  
DB 333 TMGYDKDEIYEALSESSEDTPAYNEIRNAYILK----DNKSLIKOMKQDNMVTQELDTFL 388  
QY 343 SRREPS 348  
DB 389 SQSPPT 394

RESULT 4

AAK2 HUMAN STANDARD; PRT; 552 AA.  
AC P54646; Q9HIE8; Q9UD43;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 5'-AMP-activated protein kinase, catalytic alpha-2 chain (EC 2.7.1.1-)  
DE (AMPK alpha-2 chain).  
GN PRKAA2 OR AMPK2 OR AMPK.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Heart;  
RX MEDLINE=95047501; PubMed=7959015;  
RA Aguan K., Scott J., See C.G., Sarkar N.H.;  
RT "Characterization and chromosomal localization of the human homologue  
RT of a rat AMP-activated protein kinase-encoding gene: a major  
RT regulator of lipid metabolism in mammals.";  
RL Gene 149:345-350(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=95080410; PubMed=7988703;  
RA Beri R.K., Marley A.E., See C.G., Sopwith W.F., Aguan K., Carling D.,  
RA Scott J., Carey F.;  
RT "Molecular cloning, expression and chromosomal localisation of human  
RT AMP-activated protein kinase.";  
RL FEBS Lett. 356:117-121(1994).  
RN [3]  
RP SEQUENCE OF 33-552 FROM N.A.  
RA Cobley V.;  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS  
CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES  
CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF  
CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.  
CC APPEARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE  
CC SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE  
CC DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION  
CC AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT.  
CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA  
CC NON-CATALYTIC SUBUNITS.  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SNF1 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U06454; AAA64745.1; -.  
DR EMBL; AL035705; CAC17574.1; -.  
DR PIR; S51025; S51025.

DR HSSP; Q63450; 1A06.  
DR Genew; HGNC:9377; PRKAA2.  
DR MIM; 600497; -.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR000719; Prot kinase.  
DR InterPro; IPR002290; Ser Thr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; Fatty acid biosynthesis;  
KW Phosphorylation; ATP-binding; Multigene family.  
FT DOMAIN 16 268 PROTEIN KINASE.  
FT NP\_BIND 22 30 ATP (BY SIMILARITY).  
FT BINDING 45 45 ATP (BY SIMILARITY).  
FT ACT\_SITE 139 139 BY SIMILARITY.  
FT MOD\_RES 172 172 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT CONFLICT 180 180 A -> T (IN REF. 1).  
FT CONFLICT 271 271 D -> G (IN REF. 1).  
FT CONFLICT 403 404 HL -> RQ (IN REF. 1).  
SQ SEQUENCE 552 AA; 62319 MW; C46AAFCID5104975 CRC64;

Query Match 29.2%; Score 670; DB 1; Length 552;  
Best Local Similarity 36.9%; Pred. No. 1.6e-40;  
Matches 151; Conservative 78; Mismatches 146; Indels 34; Gaps 7;

QY 1 MTKKMR-----RVGKYEVGRTIGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQ 55  
DB 1 MAEQKHGDRVKIGHYVLGVTGVTGKVGKIGEHQLTGHKAVAVKILNRQKIRSLDVVGK 60  
QY 56 IKREISIMKIVRHPNIVRLVEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYP 115  
DB 61 IKREIQNLKLFRRPHIILKLYQVISTPTDFFVMVMXYSGGELFDYICKHGRVEEMEARRLF 120  
QY 116 QQLVDVAHCHCKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNY 175  
DB 121 QQILSAVDYCHRMVVRDLKPENVLLDAHNAKIADFGLSNMMSDG-EFLRTSCGSPNY 179  
QY 176 VAPEVLSGGYDGSAAADIWSCGVILFVILAGYLPFGSDTLPGLYRKINAAEFSCPPWFSA 235  
DB 180 AAPEVISGRLYAGPEVDIWSGVILYALLCGTLFPDDEHVPTLFKKIRGGVFIPEYLNLR 239  
QY 236 EVKFLTHRLDPNPKTRIQIKKDPWRLNLYVPIRAREEE--EVNLDIDRAVFDGIE- 292  
DB 240 SVATLLMHMLQVDPLKRAITKDIREHEWFKQDLPSLYFPEDPSYDANVIDDEAVKEVCEK 299  
QY 293 ----GSYVAENVERNDEGLPMNAFEMITLSQGLNLSALFDRRDQFVKRQTRFVSRREPS 348  
DB 300 FECTSEVMNSLYSGDPQDLAVAYHLI-----IDNRRIMNQASEFYLAASSPPS 348  
QY 349 EIIANIEAVANSMSGKSHTRNF-----KTR--LEGLSSIKAGQLAV 387  
DB 349 GSFMDDSAMHIPPGLKPHPERMPPLIADSPKARCPDLALNTTKPKSLAV 397

RESULT 5

SNF1\_CANTR STANDARD; PRT; 619 AA.  
ID SNF1\_CANTR  
AC O94168;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Carbon catabolite derepressing protein kinase (EC 2.7.1.1-).  
GN SNF1.  
OS Candida tropicalis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5482;  
RN [1]  
RP SEQUENCE FROM N.A.







Db 240 SIATLLMHMLQVDPLKRAIKIDIREHEWFKQDLPSYLFEPDPSYDANVIDDEAVKEVCEK 299  
QY 293 -----GSYVAENVERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPS 348  
Db 300 PECTESEVMNSLYSGDQDQQLAVAYHLI-----IDNRRIMNQASEFYLASSPPT 348  
QY 349 EILANIEAVANSMGFKSHTRNF-----KTR--LEGLSSIKAGQLAV 387  
Db 349 GSFMDDMAMHIPPGLKPHPERMPPLIADSPKARCPDLALNTTKPKSLAV 397

RESULT 8

SNF1\_CANAL STANDARD; PRT; 620 AA.  
AC P52497; Q00309;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Carbon catabolite derepressing protein kinase (EC 2.7.1.-).  
GN SNF1.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 32354 / B-311;  
RX MEDLINE=98053924; PubMed=9393775;  
RA Pether R., Chang Y.C., Kwon-Chung K.J.;  
RT "A gene homologous to Saccharomyces cerevisiae SNF1 appears to be  
RT essential for the viability of Candida albicans.";  
RL Infect. Immun. 65:4909-4917(1997).  
RN [2]

SEQUENCE OF 7-620 FROM N.A.  
STRAIN=ATCC 32354 / B-311;  
Petter R., Kwon-Chung K.J.;  
Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ESSENTIAL FOR RELEASE FROM GLUCOSE REPRESSION. IT  
CC INTERACTS AND HAS FUNCTIONAL RELATIONSHIP TO THE REGULATORY  
CC PROTEIN SNF4. COULD PHOSPHORYLATES CAT8 (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH THE NUCLEAR MEMBRANE (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SNF1 SUBFAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; L78129; AAB48643.1; -.  
DR EMBL; L39263; AAA92456.1; -.  
DR HSSP; Q63450; 1A06.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transferase; Serine/threonine-protein kinase; ATP-binding;  
KW Phosphorylation; Carbohydrate metabolism; Nuclear protein.  
FT DOMAIN 16 29 POLY-HIS.  
FT DOMAIN 53 305 PROTEIN KINASE.  
FT NP\_BIND 59 67 ATP (BY SIMILARITY).  
FT BINDING 82 82 ATP (BY SIMILARITY).  
FT ACT\_SITE 175 175 BY SIMILARITY.  
FT MOD\_RES 208 208 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CONFLICT 7 8 PQ -> AR (IN REF. 2).

FT CONFLICT 228 234 AGPEVDV -> SSVQKIMI (IN REF. 2).  
FT CONFLICT 242 255 YVMLCGRLPFDDEF -> GMSCCVVVDYHSMTSS (IN  
REF. 2).  
FT CONFLICT 387 387 A -> R (IN REF. 2).  
FT CONFLICT 416 416 R -> A (IN REF. 2).  
FT CONFLICT 494 494 S -> L (IN REF. 2).  
SQ SEQUENCE 620 AA; 70005 MW; 1806C652B5061D2B CRC64;  
  
Query Match 28.8%; Score 661.5; DB 1; Length 620;  
Best Local Similarity 44.1%; Pred. No. 7.5e-40;  
Matches 128; Conservative 75; Mismatches 76; Indels 11; Gaps 6;  
  
QY 7 RVGKYEVGRTIGETFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIV 66  
Db 49 RIGRYQILKTLGEGSGFKVLAQHLGTGQKVALKIINRKTAKSDMQGRVEREISYLRLL 108  
QY 67 RHPNIVRLYEVLASPKIYIVLEFVTGTGGELEFDRIVHKGRLESESRKYFQQLVDAVAHCH 126  
Db 109 RHPHIIKLYDVIKSKDEIIMVIEF-AGKELEFYIVQRGKMPEDEARFFQIIAAVEYCH 167  
  
QY 127 CKGVYHRDLKPENLLDNTNGNLKVSDFGLSALPQEGVELLRTTCGTENYV-APEVLSGGQ 185  
Db 168 RHKIVHRDLKPENLLDQNLNVKIADFGLSNIMTDG-NFLKTCGSENYMPAPEVISGKL 226  
QY 186 YDGSAAADIWSCGVILFVILAGYLPFSETDPLGYRKINAAEFSCPPWFSAEVKFLIHRIL 245  
Db 227 YAGPEVDVWSAGVILYVMLCGRLPFDDEFIPALFKKISNGVYTLFNLVSAGAKHLLTRML 286  
  
QY 246 DPNPKTRIQIQIKKDPWFLR---NYV---PIRAREEEEEVNLD--IRAV 287  
Db 287 VVNPLNRITHEIMEDDWFKQDMPDYLLPPDLGSKNKSIDVDEVDVIRAL 336

RESULT 9

AAK1\_HUMAN STANDARD; PRT; 550 AA.  
AC Q13131; Q00286; Q9UNQ4;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 5'-AMP-activated protein kinase, catalytic alpha-1 chain (EC 2.7.1.-)  
DE (AMPK alpha-1 chain).  
GN PRKAA1 OR AMPK1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RA Yano K.;  
RT "Nucleotide sequence of cDNA for human AMP-activated protein kinase  
RT alpha-1.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=20499367; PubMed=11042152;  
RA Zhang Q.-H., Ye M., Wu X.-Y., Ren S.-X., Zhao M., Zhao C.-J., Fu G.,  
RA Shen Y., Fan H.-Y., Lu G., Zhong M., Xu X.-R., Han Z.-G., Zhang J.-W.,  
RA Tao J., Huang Q.-H., Zhou J., Hu G.-X., Gu J., Chen S.-J., Chen Z.;  
RT "Cloning and functional analysis of cDNAs with open reading frames for  
RT 300 previously undefined genes expressed in CD34+ hematopoietic  
RT stem/progenitor cells.";  
RL Genome Res. 10:1546-1560(2000).  
RN [3]  
RP SEQUENCE OF 27-200 FROM N.A.  
RC TISSUE=Intestine;  
RA Taboada E.N., Hickey D.A.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 294-550 FROM N.A.  
RC TISSUE=Liver;

RX MEDLINE=96132781; PubMed=8557660;  
RA Stapleton D., Mitchellhill K.I., Gao G., Widmer J., Michell B.J.,  
RA Teh T., House C.M., Fernandez C.S., Cox T., Witters L.A.,  
RA Kemp B.E.;  
RT "Mammalian AMP-activated protein kinase subfamily.";  
RL J. Biol. Chem. 271:611-614(1996).  
CC -!- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS  
CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES  
CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF  
CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.  
CC APPEARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE  
CC SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE  
CC DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION  
CC AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT (BY SIMILARITY).  
CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA  
CC NON-CATALYTIC SUBUNITS.  
CC -!- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SNF1 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB022017; BAA36547.1; --  
DR EMBL; AF100763; AAD43027.1; --  
DR EMBL; U22456; AAA64850.1; --  
DR EMBL; Y12856; CAA73361.1; --  
DR PIR; G01743; G01743.  
DR HSSP; Q63450; 1A06.  
DR Genew; HGNC:9376; PRKAA1.  
DR MIM; 602739; --  
DR GO; GO:0005622; C:intracellular; IC.  
DR GO; GO:0008189; F:apoptosis inhibitor activity; NAS.  
DR GO; GO:0004691; F:CAMP-dependent protein kinase activity; NAS.  
DR GO; GO:0042557; F:eukaryotic elongation factor-2 kinase activ. . .; ISS.  
DR GO; GO:0000187; P:activation of MAPK; NAS.  
DR GO; GO:0046318; P:negative regulation of glucosylceramide bio. . .; NAS.  
DR GO; GO:0017148; P:negative regulation of protein biosynthesis; ISS.  
DR GO; GO:0045768; P:positive regulation of anti-apoptosis; NAS.  
DR GO; GO:0045542; P:positive regulation of cholesterol biosynth. . .; NAS.  
DR GO; GO:0046321; P:positive regulation of fatty acid oxidation; ISS.  
DR GO; GO:0045722; P:positive regulation of gluconeogenesis; ISS.  
DR GO; GO:0046326; P:positive regulation of glucose import; ISS.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; ISS.  
DR GO; GO:0001666; P:response to hypoxia; NAS.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferrase; Serine/threonine-protein kinase; Fatty acid biosynthesis;  
KW Phosphorylation; ATP-binding.  
KW DOMAIN 18 270 PROTEIN KINASE.  
FT NP\_BIND 24 32 ATP (BY SIMILARITY).  
FT BINDING 47 47 ATP (BY SIMILARITY).  
FT ACT\_SITE 141 141 BY SIMILARITY.  
FT MOD\_RES 174 174 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT CONFLICT 28 28 T -> A (IN REF. 3).  
FT CONFLICT 193 193 A -> V (IN REF. 3).  
FT CONFLICT 199 199 I -> L (IN REF. 3).  
FT CONFLICT 260 260 S -> T (IN REF. 2).  
SQ SEQUENCE 550 AA; 62793 MW; 3316183D744DE325 CRC64;

Query Match 28.6%; Score 655.5; DB 1; Length 550;  
Best Local Similarity 32.0%; Pred. No. 1.7e-39;  
Matches 148; Conservative 97; Mismatches 159; Indels 59; Gaps 6;  
QY 7 RVGKYEVGRITGEGTFAKVVFARNTDTGDNVAIKIMAKSTILKNMVDQIKREISIMKIV 66  
DB 14 KIGHYILGDTLGVGTFGKVKVKGKHELTGHKVAVKILNRQKIRSLDVVGKIRREIQNLKF 73  
QY 67 RHPNIVRLYEVLASPKIYIVLEFVTGGELFDRIVHKGRLSESESRKYFQQLVDAVAHCH 126  
DB 74 RHPHIIKLYQVISTPDSIFVMMEYVSGGELFDYICKNGRLDEKESRRLLFQOILSGVDYCH 133  
QY 127 CKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGGY 186  
DB 134 RHMVYHRDLKPNVLLDAHNAKIADFGLSNMMSDG-EFLRTSCGSPNYAAPEVISGRLY 192  
QY 187 DGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHRILD 246  
DB 193 AGPEVDIWSGVIYALLCGTLPFDHDDHVTLPFKKICDGFYTPQYLNPSVISLLKHMQLQ 252  
QY 247 PNPKTRIQIGIKKDPWFLNYPVIRAREE-----EEVNLDIRAVFDGIEGSYVAE 298  
DB 253 VDPMKRASIKDIREHEWFKQDLPKYLFPEDPSYSTMDIDALKEVCEKFECESEEVLS 312  
QY 299 NVERNDEGLMNAFEMITLSQGL-----NLSALFDRRQDFVKR 337  
DB 313 LYNRHQDPLAV-AVHLIIDNRIMNEAKDFYLATSPDSDFLDDHLLTRPHERVPFLVA 371  
QY 338 QT-----RFSRREPSEIIANIEAVANSMGFKSHTRN- 369  
DB 372 ETTPRARHTLDELNPQSKHQGVKAKWHLGIRSQSRPNIDMAEVCRAIKQLDYEWKVNVP 431  
QY 370 FKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVKAAGETLE 412  
DB 432 YLVRVRKKNPVTSTYSKMSLQLYQVDSRTYLLDFRSIDDEITE 474  
RESULT 10  
AAKL\_RAT  
ID AAKL\_RAT STANDARD; PRT; 548 AA.  
AC P54645;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 5'-AMP-activated protein kinase, catalytic alpha-1 chain (BC 2.7.1.-)  
DE (AMPK alpha-1 chain).  
GN PRKAA1 OR AMPK1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Liver;  
RX MEDLINE=96132781; PubMed=8557660;  
RA Stapleton D., Mitchellhill K.I., Gao G., Widmer J., Michell B.J.,  
RA Teh T., House C.M., Fernandez C.S., Cox T., Witters L.A.,  
RA Kemp B.E.;  
RT "Mammalian AMP-activated protein kinase subfamily.";  
RL J. Biol. Chem. 271:611-614(1996).  
RN [2]  
RP PARTIAL SEQUENCE.  
RC STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Liver;  
RX MEDLINE=95050763; PubMed=7961907;  
RA Stapleton D., Gao G., Michell B.J., Widmer J., Mitchell K.,  
RA Teh T., House C.M., Witters L.A., Kemp B.E.;  
RT "Mammalian 5'-AMP-activated protein kinase non-catalytic subunits are  
RT homologs of proteins that interact with yeast Snf1 protein kinase.";  
RL J. Biol. Chem. 269:29343-29346(1994).  
CC -!- FUNCTION: RESPONSIBLE FOR THE REGULATION OF FATTY ACID SYNTHESIS  
CC BY PHOSPHORYLATION OF ACETYL-COA CARBOXYLASE. IT ALSO REGULATES  
CC CHOLESTEROL SYNTHESIS VIA PHOSPHORYLATION AND INACTIVATION OF  
CC HORMONE-SENSITIVE LIPASE AND HYDROXYMETHYLGLUTARYL-COA REDUCTASE.  
CC APPEARS TO ACT AS A METABOLIC STRESS-SENSING PROTEIN KINASE  
CC SWITCHING OFF BIOSYNTHETIC PATHWAYS WHEN CELLULAR ATP LEVELS ARE  
CC DEPLETED AND WHEN 5'-AMP RISES IN RESPONSE TO FUEL LIMITATION  
CC AND/OR HYPOXIA. THIS IS A CATALYTIC SUBUNIT (BY SIMILARITY).  
CC -!- SUBUNIT: HETEROTRIMER OF A CATALYTIC SUBUNIT, A BETA AND A GAMMA  
CC NON-CATALYTIC SUBUNITS.  
CC -!- PTM: AUTOPHOSPHORYLATED (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
CC SNF1 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB022017; BAA36547.1; --  
DR EMBL; AF100763; AAD43027.1; --  
DR EMBL; U22456; AAA64850.1; --  
DR EMBL; Y12856; CAA73361.1; --  
DR PIR; G01743; G01743.  
DR HSSP; Q63450; 1A06.  
DR Genew; HGNC:9376; PRKAA1.  
DR MIM; 602739; --  
DR GO; GO:0005622; C:intracellular; IC.  
DR GO; GO:0008189; F:apoptosis inhibitor activity; NAS.  
DR GO; GO:0004691; F:CAMP-dependent protein kinase activity; NAS.  
DR GO; GO:0042557; F:eukaryotic elongation factor-2 kinase activ. . .; ISS.  
DR GO; GO:0000187; P:activation of MAPK; NAS.  
DR GO; GO:0046318; P:negative regulation of glucosylceramide bio. . .; NAS.  
DR GO; GO:0017148; P:negative regulation of protein biosynthesis; ISS.  
DR GO; GO:0045768; P:positive regulation of anti-apoptosis; NAS.  
DR GO; GO:0045542; P:positive regulation of cholesterol biosynth. . .; NAS.  
DR GO; GO:0046321; P:positive regulation of fatty acid oxidation; ISS.  
DR GO; GO:0045722; P:positive regulation of gluconeogenesis; ISS.  
DR GO; GO:0046326; P:positive regulation of glucose import; ISS.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; ISS.  
DR GO; GO:0001666; P:response to hypoxia; NAS.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transferrase; Serine/threonine-protein kinase; Fatty acid biosynthesis;  
KW Phosphorylation; ATP-binding.  
KW DOMAIN 18 270 PROTEIN KINASE.  
FT NP\_BIND 24 32 ATP (BY SIMILARITY).  
FT BINDING 47 47 ATP (BY SIMILARITY).  
FT ACT\_SITE 141 141 BY SIMILARITY.  
FT MOD\_RES 174 174 PHOSPHORYLATION (AUTO-) (POTENTIAL).  
FT CONFLICT 28 28 T -> A (IN REF. 3).  
FT CONFLICT 193 193 A -> V (IN REF. 3).  
FT CONFLICT 199 199 I -> L (IN REF. 3).  
FT CONFLICT 260 260 S -> T (IN REF. 2).  
SQ SEQUENCE 550 AA; 62793 MW; 3316183D744DE325 CRC64;









```
RESULT 13
YNA3 CAEEL          STANDARD;          PRT;          622 AA.
AC P45894;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative serine/threonine-protein kinase PAR2.3 in chromosome III
DE (EC 2.7.1.-).
GN PAR2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laistner N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U00025; AAA50618.1; -.
CC PIR; S44859; S44859.
CC HSSP; Q63450; 1A06.
CC WormPep; PAR2.3; C600847.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00220; S_TKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
CC Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding.
KW DOMAIN 24 276 PROTEIN KINASE.
FT NP_BIND 30 38 ATP (BY SIMILARITY).
FT BINDING 53 53 ATP (BY SIMILARITY).
FT ACT_SITE 147 147 BY SIMILARITY.
SQ SEQUENCE 622 AA; 69515 MW; DD5FB1555004B246 CRC64;
```

```
Query Match          25.8%; Score 592.5; DB 1; Length 622;
Best Local Similarity 40.7%; Pred. No. 6.4e-35;
Matches 116; Conservative 62; Mismatches 102; Indels 5; Gaps 2;

QY 7 RVGKVEVGRITIGETFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIV 66
Db 20 KIGNFVIKETIGKAFGAVKRGTHIQTDYDAIKILNRGMKGLGTVNKTRNEIDNLQKL 79
QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYPQQLVDAVACH 126
Db 80 THEHITLFRVISTPDSIFLVMELVSGGELFSYTRKGAIPRESRRYFQQLISGVSYCH 139
```

```
QY 127 CKGVYHRDLKPENLLDITNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGOY 186
Db 140 NHMIVHRDLKPENLLDANKNIKIADFGLSNYMTDG-DLLSTACGSPNYAAPELISNKLY 198
QY 187 DGSAAADIWSCGVILFVILAGYLPSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILD 246
Db 199 VGPEVDLWSCGVILYAMLCGTLFPDDQNVPTLFAKIKSGRYTPVPSMEKQAADLISTMLQ 258
QY 247 PNPKTRIQIQGIKKDPWFLRN---YVPIRAREEEVNLDIRAV 287
Db 259 VDFVKRADVKRIVNHSWFRIDLPLYLFPCEENESSIVDIDVVQSV 303

RESULT 14
KKK1_YEAST
ID KKK1_YEAST          STANDARD;          PRT;          1518 AA.
AC P34244;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Probable serine/threonine-protein kinase YKL101W (EC 2.7.1.-).
GN YKL101W OR YKL453.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94078677; PubMed=8256524;
RA Pallier C., Valens M., Puzos V., Fukuhara H., Cheret G., Sor F.,
RA Bolotin-Fukuhara M.;
RT "DNA sequence analysis of a 17 kb fragment of yeast chromosome XI
RT physically localizes the MRB1 gene and reveals eight new open reading
RT frames, including a homologue of the KIN1/KIN2 and SNF1 protein
RT kinases.";
RL Yeast 9:1149-1155(1993).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC NIM1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X71133; CAA50456.1; -.
CC EMBL; Z28101; CAA81941.1; -.
CC PIR; S37928; S37928.
CC HSSP; Q63450; 1A06.
CC SGD; S0001584; HSL1.
CC GO; GO:0005935; C:bud neck; IDA.
CC GO; GO:0005940; C:septin ring; IDA.
CC GO; GO:0000086; P:G2/M transition of mitotic cell cycle; IGI.
CC GO; GO:0006468; P:protein amino acid phosphorylation; IDA.
CC GO; GO:0000074; P:regulation of cell cycle; IMP.
CC GO; GO:0000135; P:septin checkpoint; IGI.
CC InterPro; IPR000719; Prot_kinase.
CC InterPro; IPR002290; Ser_thr_pkinase.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Prot_kinase; 1.
CC SMART; SM00220; S_TKC; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW ATP-binding.
KW DOMAIN 81 369 PROTEIN KINASE.
FT
```

```
FT NP_BIND 87 95 ATP (BY SIMILARITY).
FT BINDING 110 110 ATP (BY SIMILARITY).
FT ACT_SITE 239 239 BY SIMILARITY.
SQ SEQUENCE 1518 AA; 169592 MW; 803F84F7531241DD CRC64;

Query Match 25.7%; Score 589.5; DB 1; Length 1518;
Best Local Similarity 30.7%; Pred. No. 3.2e-34;
Matches 138; Conservative 90; Mismatches 135; Indels 87; Gaps 10;

QY 2 TKQVRR-----VGKVEVGRITGEGTFAKVKFARNVTDGDNVAIKIMAKSTIL----- 48
Db 67 TKSSKRKSRDITVGPWKLGKTLGKSSGRRVRLAKNMTGQLAAIKIVPKKAFVHCSNNGT 126
QY 49 -----KNRMVDQ-----IKREISIMKIVRHPENIVLYEVLASP 81
Db 127 VPNSYSSSMVTSNVSSPSIASREHSNHSQTNPYGIEREIVIMKLISHTNVMALEFVWENK 186
QY 82 SKIYIVLEFVTGGELFDRIVHKRLEESERKYPQQLVDAVAHCHCKGVYHRDLKPENLL 141
Db 187 SELYLVLVEYVDGGELFDLYLVSKGLPEREAIHYFKQIVEGVSYCHSFNICHRLDKPENLL 246
QY 142 LD-TNGNLKVSDFGLSALPQEGVELLRTTCGTPNRYVAPEVLGQGYDGSAAIWSCGVIL 200
Db 247 LDKNRRRIKIADFGMAALELPN-KLLKTSQSPHYASPEIVMGRPYHGGPSDWSGIVL 305
QY 201 FVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQIGIKK 260
Db 306 FALLTGHLPFNDNDNIKLLKLVQSGKYQMPNSLSSEARDLISKILVIDEKRITTEILK 365
QY 261 DPWFR-----LNVVPIRAREE-----EEVNLDDI 284
Db 366 HPLIKKYDDLVPNKVLRKMRKDNMARGKNSDLHLNNVSPSIVTLHSEKGEIDESILRSL 425
QY 285 RAVFDGIEGSYVAENVERN--DEGLMMNAFEMITLSQGLNLSALFRRQDFVKRQTRFV 342
Db 426 QILWHGVSRELITAKLLQKPMSEKLFYSLLLQYKQKHSISLSSSEN-----KKSATES 480
QY 343 SRREPSEIIANIEAVANSMSGFKSHTRNFKT 372
Db 481 SVNEPR--LEYASKTANTGLRSENNDVKT 508

RESULT 15
MRK4 HUMAN
ID MRK4 HUMAN STANDARD; PRT; 752 AA.
AC Q96L34; Q96JG7; Q9BYD8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE MAP/microtubule affinity-regulating kinase 4 (EC 2.7.1.27)
DE (MAP/microtubule affinity-regulating kinase like 1).
GN MARK4 OR MARKL1 OR KIAA1860.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Drewes G., Mandelkow E.M.;
RT "MARK4, homologue of MARK1, MARK2 and MARK3.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=21226021; PubMed=11326310;
RA Kato T., Satoh S., Okabe H., Kitahara O., Ono K., Kihara C.,
RA Tanaka T., Tsunoda T., Yamaoka Y., Nakamura Y., Furukawa Y.;
RT "Isolation of a novel human gene, MARKL1, homologous to MARK3 and its
RT involvement in hepatocellular carcinogenesis.";
RL Neoplasia 3:4-9(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
```

```
RC TISSUE=Brain;
RA Beghini A., Magnani I., Roversi G., Piepoli T., DiTerlizzi S.,
RA Pollo B., Conti A.M.F., Cowell J.K., Finocchiaro G., Larizza L.;
RT "Neural progenitor-restricted isoform of MARKL1 gene is upregulated by
RT 19q13 amplification in human glioblastoma.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Thyroid;
RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q96L34-1; Sequence=Displayed;
CC Name=2; Synonyms=MARKL1S;
CC IsoId=Q96L34-2; Sequence=VSP_004946;
CC -!- TISSUE SPECIFICITY: Ubiquitous. Isoform 2 is brain-specific.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. MARK
CC SUBFAMILY.
CC -!- SIMILARITY: Contains 1 kinase-associated (KAI) domain.
CC -!- SIMILARITY: Contains 1 UBA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AY057448; AAL23683.1; -
EMBL; AB049127; BAB39380.1; -
EMBL; AY120867; AAM55491.1; -
EMBL; AK075272; BAC11510.1; -
EMBL; AB058763; BAB47489.1; ALT_INIT.
HSSP; Q63450; 1A06.
Genew; HGNC:13538; MARK4.
MIM; 606495; -
InterPro; IPR001772; Kinase Cterm.
InterPro; IPR000719; Prot_kinase.
InterPro; IPR002290; Ser_thr_kinase.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR000449; UBA_domain.
Pfam; PF02149; KAI; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF00627; UBA; 1.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Prot_kinase; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
PROSITE; PS50030; UBA; 1.
Transferase; Serine/threonine-protein kinase; ATP-binding;
Alternative splicing.
DOMAIN 59 310 PROTEIN KINASE.
DOMAIN 324 368 UBA.
DOMAIN 703 752 KAI.
```

```
FT NP BIND      65 73      ATP (BY SIMILARITY).
FT BINDING      88 88      ATP (BY SIMILARITY).
FT ACT SITE     181 181     BY SIMILARITY.
FT VARSPLIC     628 752     ADEPERIGGPEVTSCHLPWDQDETAPRLRFPWSVKLTSSR
FT PPEALMAALRQATAAARCRQPOPFLLACLGAGGPEPL
FT SHFEVEVCQLPRPGRLGVLFRRVAGTALAFRTLVTRISNDL
FT EL -> TLDPSKRONRNCVSGASLPQGSKIRSQTNLRESG
FT DLRSQVAIYLGIRKPPPCSDSPGV (in isoform
FT 2).
FT /FTid=VSP 004946.
FT CONFLICT     70 70      F -> S (IN REF. 2).
SQ SEQUENCE     752 AA; 82519 MW; 4B430FFD2B150E7A CRC64;

Query Match      25.6%; Score 586; DB 1; Length 752;
Best Local Similarity 40.6%; Pred. No. 2.3e-34;
Matches 121; Conservative 60; Mismatches 113; Indels 4; Gaps 3;

QY 8 VGKYEVCRTIGETFAKVKFARNVTDGDNVAIKIMAKSTILKNRMVDOIKREISIMKIVR 67
Db 56 VGNTRLRLRTIGKGNFAKVKLARHILTGREVAIKIIDK-TQLNPSSLQKLFREVRIMKGLN 114
QY 68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESESRKYFQQLVDVAHCHC 127
Db 115 HPNIVKLEFEVIEETKTLVLMVEYASAGEVFDYLVSHGRMKEKEARAKFRQIVSAVHYCHQ 174
QY 128 KGVYHRDLKPENLLDNTGNLKVSDPGLSALPOEGVELLRTTCGTPNYVAPEVLSGQGYD 187
Db 175 KNIVHRDLKAENLLDAAENIKIADFGFSNEFTLGSK-LDTFCGSPPPYAAPLEFGKKYD 233
QY 188 GSAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAHFSCPPWFSAEVKFLIHRILDP 247
Db 234 GPEVDIWSLGVILYTLVSGSLPFDGHNKELRERVLRGKYRVPFYMSTDCESILRRFLVL 293
QY 248 NPKTRIQQGIKKDPWFRNLNY--VPIRAREEEVNLDDIRAVFDGIEGSYVAENVERN 303
Db 294 NPAKRCTLEQIMKDKWINIGYEGELKPYTEPEEDFGDTKRIEVMVGMGYTREIKES 351
```

Search completed: October 20, 2003, 16:51:08  
Job time : 17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:48:42 ; Search time 19 Seconds  
(without alignments)  
2257.433 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTCKMRRVGKYEVRITIGEG.....IINRATEGIPKSEILRTITP 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
1	1484.5	64.7	445	2 T09903	serine/threonine-s
2	1381.5	60.2	480	2 A86427	probable serine/th
3	1333	58.1	440	2 T14735	probable serine/th
4	1314	57.3	441	2 C84667	probable protein k
5	1313	57.3	440	2 T14736	probable serine/th
6	1234.5	53.8	453	2 G86141	protein T25K16.13
7	1040	45.4	441	2 E85362	hypothetical prote
8	1009.5	44.0	445	2 T50802	serine/threonine p
9	1006.5	43.9	456	2 T48478	serine/threonine p
10	999.5	43.6	421	2 E96522	hypothetical prote
11	983.5	42.9	489	2 T04862	probable serine/th
12	979.5	42.7	421	2 T48202	protein kinase AK2
13	978	42.7	520	2 G86414	probable protein k
14	968.5	42.2	461	2 T14822	probable serine/th
15	935	40.8	502	2 T02306	probable protein k
16	924	40.3	469	2 B84644	probable protein k
17	900.5	39.3	439	2 T03444	protein kinase hom
18	879	38.3	435	2 E84707	probable protein k
19	869.5	37.9	426	2 C71408	probable protein k
20	869	37.9	442	2 T48203	hypothetical prote
21	845	36.9	527	2 A53467	protein kinase SNF
22	808.5	35.3	431	2 T02496	probable protein k
23	718	31.3	472	2 B90100	SNF-related kinase
24	705	30.7	511	1 A56009	serine/threonine-s
25	701.5	30.6	576	2 T41587	probable carbon ca
26	700.5	30.5	512	1 JC1446	serine/threonine-s
27	686	29.9	504	2 T10449	probable serine/th
28	677	29.5	512	2 T52633	serine/threonine-s
29	676	29.5	504	2 T07415	probable serine/th

30	673	29.4	473	1 S59941	serine/threonine-s
31	670	29.2	552	1 S51025	lhydroxymethylglut
32	666	29.0	512	2 T07788	probable serine/th
33	664	29.0	562	2 T29858	hypothetical prote
34	663	28.9	633	1 A26030	serine/threonine-s
35	662	28.9	552	1 A53621	lhydroxymethylglut
36	648	28.3	602	2 S72513	FOG2 protein - yea
37	638	27.8	510	2 T04145	serine/threonine p
38	634	27.6	513	1 S60303	serine/threonine-s
39	624.5	27.2	513	1 S60304	serine/threonine-s
40	614.5	26.8	651	2 S52244	p69Eg3 protein - A
41	610	26.6	502	1 A41361	serine/threonine-s
42	607	26.5	887	2 T20941	hypothetical prote
43	598	26.1	1142	2 S59359	GIN4 protein - yea
44	597.5	26.1	1192	2 T18611	probable serine/th
45	597.5	26.1	1246	2 G89287	protein H39E23.1 [

ALIGNMENTS

RESULT 1

T09903  
serine/threonine-specific protein kinase homolog T22A6.230 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 15-Oct-1999  
C:Accession: T09903  
R:Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May  
submitted to the Protein Sequence Database, June 1999  
A:Reference number: Z16896  
A:Accession: T09903  
A:Molecule type: DNA  
A:Residues: 1-445 <BEV>  
A:Cross-references: EMBL:AL078637; GSPDB:GN00062; ATSP:T22A6.230  
A:Experimental source: cultivar Columbia; BAC clone T22A6  
C:Genetics:  
A:Gene: ATSP:T22A6.230  
A:Map position: 4  
A:Introns: 53/3; 74/3; 98/3; 134/3; 159/3; 177/3; 219/3; 249/3; 290/3; 329/3; 368/3; 387  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolc

Query Match 64.7%; Score 1484.5; DB 2; Length 445;  
Best Local Similarity 62.3%; Pred. No. 2.4e-56;  
Matches 273; Conservative 84; Mismatches 80; Indels 1; Gaps 1;

QY	5	MRRVGKYEVRRTIGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMK 64
DB	3	VRKVGKYEVRRTIGEGTFAKVKFAQNTETGESVAMKIVDRSTIIKRKMVDQIKREISIMK 62
QY	65	IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELEFDRIVHKGRLSESRKYFQQLVDAVAH 124
DB	63	LVRHPCVRLYEVLASRTKIYIILEYITGGELEFDRIVHKGRLSESRKYFQQLIDVDY 122
QY	125	CHCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNVVAPEVLSGQ 184
DB	123	CHSKGVYHRDLKPNLLDLSQGNLKISDFGLSALPEQGVTLTKTCGTPNVVAPEVLSHK 182
QY	185	GYDGAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHR 244
DB	183	GYGAVADIWSCGVILYVLMAGYLPFDEMDLPFLYSKIDKAEFSCPSYFALGAKSLINR 242
QY	245	LDNPKTRIQIOGIKKDPFRLNVYPVIRAREEEVNLDIDRAVFDGIEGYSVAENVERND 304
DB	243	LDNPKTRITIAEIRKDEWFLKDYTPVQLIDYEHVNLDDVYAAFDDEEQTQAQDTR-D 301
QY	305	EGPLMNAFEMITLSQGLNLSALFDRRQDFVKEQTRFVSRREPSEIIANIEAVANSMGFK 364
DB	302	TGPLTLNADFLLIILSQGLNLSALFDRRQDFVKEQTRFVSRREPSEIIANIEAVANSMGFK 361
QY	365	SHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPELPMVDVRKAAGETLEVHKFKLCSKL 424
DB	362	THIRNYKMRVEGLSANKTSHFVILEVFKVAPSIILMVDIQNAAGDAEEYLFYKTFCSKL 421





A:Map position: 2  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 57.3%; Score 1314; DB 2; Length 441;  
Best Local Similarity 57.7%; Pred. No. 3.8e-49;  
Matches 248; Conservative 81; Mismatches 95; Indels 6; Gaps 3;

QY 6 RRVGKVEVGRITIGETFAKVKFARNSTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKI 65  
DB 9 RRVGKVEVGRITIGETFAKVKFARNSTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKI 68

QY 66 VRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVDAVHC 125  
DB 69 IKHPNVQLYEVMAKTKFIILEYVTGGELFDKIVNDGRMKDEARRVFQQLIHAVDYC 128

QY 126 HCKGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAEVL 183  
DB 129 HSRGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAEVL 188

QY 184 QGYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHR 243  
DB 189 RGYDGATADWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHR 248

QY 244 ILDPNPKTRIQIQTGKDPFRLNVPYPIRAREEEVNLDIDIRAVFDGIEGYSVAENV 303  
DB 249 ILDPNPKTRIQIQTGKDPFRLNVPYPIRAREEEVNLDIDIRAVFDGIEGYSVAENV 305

QY 304 DEGLMNAFEMITLSQGLNLSALFDRRQDFVVKRQTRFVSRRPSEIIANIEAVANSMGF 363  
DB 306 EEQPAANAFEMITLSQGLNLSALFDRRQDFVVKRQTRFVSRRPSEIIANIEAVANSMGF 364

QY 364 KSHTRNFKTRLEGLSSIKAGQAVVIEYVAPSLFMDVVRKAAGETLEYHKFKYKLSK 423  
DB 365 DVQKKNYKRLNENKAGRKGNLNVATEIFQVAPSLHVMQVSKSGDTLEFHKFKYKLSK 424

QY 424 LENIWRATE 433  
DB 425 LEQVVTNNE 434

RESULT 5  
T14736  
probable serine/threonine kinase (EC 2.7.1.1-) SNFL2 - sorghum  
N:Alternate names: SNF1 protein kinase homolog SNFL2  
C:Species: Sorghum bicolor (sorghum)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C:Accession: T14736  
R:Annen, F.; Stockhaus, J.  
Plant Mol. Biol. 36, 529-539, 1998  
A:Title: Characterization of a Sorghum bicolor gene family encoding putative protein kinases  
A:Reference number: Z18177; MUID:98145442; PMID:9484448  
A:Accession: T14736  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-440 <ANN>  
A:Cross-references: EMBL:Y12465; NID:g2632253; PIDN:CAA73068.1; PID:g2632254  
A:Experimental source: cultivar TX 430  
C:Genetics:  
A:Gene: SNFL2  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C:Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 57.3%; Score 1313; DB 2; Length 440;  
Best Local Similarity 58.3%; Pred. No. 4.2e-49;  
Matches 253; Conservative 76; Mismatches 99; Indels 6; Gaps 3;

QY 2 TKQMRVGVKVEVGRITIGETFAKVKFARNSTDTGDNVAIKMAKSTILKNRMVDQIKREIS 61  
DB 4 TKVRRVGVKVEVGRITIGETFAKVKFARNSTDTGDNVAIKMAKSTILKNRMVDQIKREIS 63

QY 62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVDA 121  
DB 64 TMKLIKHPNVVRIYEVMSKTIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVDA 123

QY 122 VAHCHCKGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAE 179  
DB 124 VDCHSRGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAE 183

QY 180 VLSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVK 239  
DB 184 VLEDQGYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVK 243

QY 240 LTHRILDPNPKTRIQIQTGKDPFRLNVPYPIRAREEEVNLDIDIRAVFDGIEGYSVAEN 299  
DB 244 LTHRILDPNPKTRIQIQTGKDPFRLNVPYPIRAREEEVNLDIDIRAVFDGIEGYSVAEN 300

QY 300 VERNDEGLMNAFEMITLSQGLNLSALFDRRQDFVVKRQTRFVSRRPSEIIANIEAVAN 359  
DB 301 TEKKEEPVVLNAPFELISRSAGNLGNLFDSQEF-KRETRFTSKCPKEIVRKIEAAK 359

QY 360 SMGFKSHTRNFKTRLEGLSSIKAGQAVVIEYVAPSLFMDVVRKAAGETLEYHKFKYK 419  
DB 360 PLGFGVQKKNYKRLNENKAGRKGNLNVATEIFQVAPSLHVMQVSKSGDTLEFHKFKYK 419

QY 420 LCKLENIIWRATE 433  
DB 420 LSKTLKDVVWKSSED 433

RESULT 6  
G86141  
protein T25K16.13 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: G86141  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, K.; Creasy, T.H.; Dewar, K.;  
Ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G86141  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-453 <STO>  
A:Cross-references: GB:AE005172; NID:g6715641; PIDN:AAF26468.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: T25K16.13  
A:Map position: 1  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match 53.8%; Score 1234.5; DB 2; Length 453;  
Best Local Similarity 53.9%; Pred. No. 8.9e-46;  
Matches 233; Conservative 86; Mismatches 100; Indels 13; Gaps 5;

QY 7 RVGKVEVGRITIGETFAKVKFARNSTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKI 66  
DB 15 RVGNVEMGRITIGESFAKVKYAKNTVTGDAQAIKLDREKVFHKKVVEQLKREISTMKLI 74

QY 67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKVFQQLVDAVACH 126  
DB 75 KHPNVVEIIEVMASKTIYIVLELVNGGELFDKIAQQGRLEKDEARRVFQQLINAVDYCH 134

QY 127 CKGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAEVL 184  
DB 135 SRGVYHRDLKPNLLDITNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAEVL 194

QY 185 GYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKIN--AAEFSCPPWFSAEVKFLI 242  
DB 195 GYDGAADVWSCGVILFVILAGYLPFSETDPLGLYRKIN--AAEFSCPPWFSAEVKFLI 254

QY	243	RILDPNPKTRIQIGIKKDPWFLRNVVPIR-AREEEVNLDIDRAVFDGIEGSYVAENVE	301
		: :     :     :     :     :     :	
Db	255	RILEPNPITRISIAELLEDEWFKKGYKPPSPDQDDDEDITIDVDAPFNSKECLVTEKKE	314
QY	302	RNDEGPLMNAFEMITLSQGLNLSALEDRRQDFVKRQTRFVSRREPSEIIANIEAVANSM	361
		: :             : :     :     :     :     :     :	
Db	315	K- - - -PVSMNAFELISSSEFSLNLEFQKQALVKKETRFTSQRSASEIMSKMEETAKPL	370
QY	362	GFKSHTRNFKTRLEGLSSIKAGOLAVVIBIYEVPAPSLFMVDVRKAAGETLEYHK- - - -FY	417
		:     :     :     :     :     :     :     :     :	
Db	371	GFNVKDNKYIKMKGDKSGRKGQLSVATEVFVEVAPSLHVVELRKTGGDTLEFHFKVDSFY	430
QY	418	KKLCSKLENIIW	429
		: : :	
Db	431	KNFSSGLKDVVW	442

RESULT 7  
E85362  
hypothetical protein AT4g30960 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
C;Accession: E85362  
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A;Reference number: A85001; MUID:20083488; PMID:10617198  
A;Accession: E85362  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-441 <STO>  
A;Cross-references: GB:NC\_001268; NID:g7269998; PIDN:CAB79814.1; GSPDB:GN00140  
C;Genetics:  
A;Gene: AT4g30960  
A;Map position: 4  
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

Query Match	45.4%;	Score 1040;	DB 2;	Length 441;
Best Local Similarity	49.0%;	Pred. No. 1.4e-37;		
Matches 217;	Conservative 73;	Mismatches 109;	Indels 44;	Gaps 10;

  

QY	9	GKYEVRGTIGEGTFAKVVFARN	TDGDNVAIKIMAKSTILK	NRMVDQIKREISIMKIVRH	68
Db	22	GRYELGRLLGHGTFAKVYHARNI	QTGKSVAMKVWGKEKVVKVGM	VDQIKREISVMRMVKH	81
QY	69	PNIVRLYEVLASPSKIYIVLEF	VTGGELFDRIVHKGRLEESER	KYFQQLVDVAHCHCK	128
Db	82	PNIVELHEVMASKSKIYFAMEL	YRGGELFAKVA-KGRLEDVAR	YFQQLISAVDFCHSR	140
QY	129	GVYHRDLKPENLLLD	TGNLKVSDPGLSA---	LPQEGVELLRTTCGTPNYVAPEVLSGQ	184
Db	141	GVYHRDLKPENLLLD	DEGNLKVTD	PGLSAFTEHLKQDG--LLHTTCGTPAYVAPEVILKK	198
QY	185	GVDGSAADIWSCGVILFVILAG	YLPFSETDLPGLYRKINAA	BFPSCPPWFSAEVKFLIHRI	244
Db	199	GVDGAKADLWSCGVILFVLLAG	YLPFQDDNLVMYRKIYRG	FKPCPGLWSSDARRLVTKL	258
QY	245	LDPNPKTRIQIQGIKKDPWF---	RLNYVPIRA---	REEEVENLDDIRAVFDGIEGSYV	296
Db	259	LDPNENTRITIEKVMDSPWF	KKQATRSRNEPVAATITTT	TEEDVDF-----	303
QY	297	AENYERNDEGPLMWNAFEMIT	LSQGLNLSALFDRRQDFV	KRQTRFVSRREPSEIIANIEA	356
Db	304	--LVHKSKEETETLNAFHII	ALSEGFDLSPLFEKKKEEK	REMRFATSRPASSVISSLEE	361
QY	357	VANSMGFKSHTR--NFKTR	LEGLSSIKAGQLAVVIEI	YEVAPSLFMVDVRKAAGETLEYH	414
Db	362	AAR-VGNKFDVRKSES	RVRIEGKQNGRKGKLA	VEAEIFAVAPSFVVVEVKKHGDTLEYN	420
QY	415	KFYKLCSS----	KLENIIRAT	432	
Db	421	NF-----CSTALRP	ALKDIFWTST	439	

RESULT 8

T50802

serine/threonine protein kinase-like protein - Arabidopsis thaliana

N/Alternate names: protein T30N20\_200

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 18-Aug-2000

C/Accession: T50802

R/Byan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, July 2000

A/Reference number: Z25240

A/Accession: T50802

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-445 <BEV>

A/Cross-references: EMBL:AL365234

A/Experimental source: cultivar Columbia; BAC clone T30N20

C/Genetics:

A/Map position: 5

A/Note: T30N20\_200

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 44.0%; Score 1009.5; DB 2; Length 445;

Best Local Similarity 47.8%; Pred. No. 2.8e-36;

Matches 213; Conservative 81; Mismatches 111; Indels 41; Gaps 11;

QY 6 RRV--GKYEVRTIGEGTFAKVKFARNRDTGDNVAIKIMAKSTILKN-RMVDQIKREISI 62

DB 5 RRVLFQKYMGRLLGKGTFAKVVYKKEITGGECVAIKVINKDQVMKRPFGMEQIKREISI 64

QY 63 MKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVDVAV 122

DB 65 MKLVRHPNIVELKEVMATKTKIFFVMEFVKGGEFCKI-SKGLHEDAARRYFQQLISAV 123

QY 123 AHCHCKGVVHRDLKPENLLDTNGNLKVSDFGLSALPQEGVE--LLRTTCGTPNYVAPEV 180

DB 124 DYCHSRGVSHRDLKPENLLDENGDLKISDFGLSALPEQILQDGLLHTQCGTPAYVAPEV 183

QY 181 LSGQGYDGSAAIDWSCGVILFVILAGVLPFFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240

DB 184 LKKGYDGAKADIWSCGVLYVLLAGCLPFDQDENIMMYRKIFRADEFFFPWFSPPEARLL 243

QY 241 IHRILDPNPKTRIQIGIKKDPWFLNYYPIRA-----REEEVNLDDIRA 286

DB 244 ISKLLVDPDRRISIPAIMRTPWLNRKNFTPLAFKIDEPICSSQSKNNEEEEDGD--- 299

QY 287 VFDGIEGYSYVAENVERNDEGPLMMNAFEMI-TLSQGLNLSALFDRRQDFVKQTRFVSRR 345

DB 300 -----CEN-QTEPISPKFFNAFEFISMSGGFDLSLFSFKR---KVQSVFTSRS 345

QY 346 EPSEIIANIEAVANSMGFK-SHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVR 404

DB 346 SATEVMEKIEITVTKEMMKVKRTKDFKVKMEGTEGRKGRLSMTAEVFEVVAPEISVVEFC 405

QY 405 KAAGETLEYHKFY-KKLCCKLENIW 429

DB 406 KSAGDTLEYDREVRPALNDIVW 431

```

RESULT 9
T48478
serine/threonine protein kinase-like protein - Arabidopsis thaliana
N;Alternate names: protein T28J14.10
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 02-Sep-2000
C;Accession: T48478
R;Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Newes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A;Reference number: Z24493
A;Accession: T48478
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-456 <BEV>

```







Db 200 KGYDAAKVDIWSCGVILFVLMAGYLPFFHDRNVNMYKKIYRGEFRCRPFSTELTRLLSK 259

Qy 244 ILDPNPKTRIQIGIKKDPWFLNLYVPIR-----AREEEVNLDLDIRAVFDGIEG 293

Db 260 LLETNPEKRTTFPEIMENSWFKGFKHKKFYVDDDKLNCNVVDDDELESDSVESDRDASA 319

Qy 294 SYVAENVE-RNDEG-----PLMMNAFEMITLSQGLNLSALFDRRQDFVKKQTRFVSRRRPS 348

Db 320 ESEIEYLEPRRRVGGULPRPASLNADFIIISFSGFDLSGLFDDDDGE---GSRFVSGAPVS 375

Qy 349 EIIANIEAVANSMGFKSHTRNFKTRLEGSLSSIKAGQLAVVIEIYEVAPSLFMDVVRKAAG 408

Db 376 KIISKLEIEIAKVVSFTVRKKDCRVSLGSRQGVKGPLTIAAEIFELTPSLVWVEVKKKG 435

Qy 409 ETLEYHKF-YKKLCSKLENI 427

Db 436 DKTEYEDFCNNELKPKLQNL 455

RESULT 12

T48202

protein kinase AK22 (EC 2.7.1.1-) - Arabidopsis thaliana

N;Alternate names: protein T20L15.80

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence revision 20-Apr-2000 #text\_change 17-Nov-2000

C;Accession: T48202; S66335; S58261

R;Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew

submitted to the Protein Sequence Database, March 2000

A;Reference number: Z24488

A;Accession: T48202

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-421 <BEV>

A;Cross-references: EMBL:AL162351

A;Experimental source: cultivar Columbia; BAC clone T20L15

R;Thuemmler, F.; Kirchner, M.; Teuber, R.; Ditttrich, P.

Plant Mol. Biol. 29, 551-565, 1995

A;Title: Differential accumulation of the transcripts of 22 novel protein kinase genes in

A;Reference number: S66314; MUID:96123233; PMID:8534852

A;Accession: S66335

A;Molecule type: DNA

A;Residues: 136-193 <THU>

A;Cross-references: EMBL:X86967; NID:g928911; PIDN:CAA60530.1; PID:g928912

C;Genetics:

A;Gene: AK22

A;Map position: 5

A;Note: T20L15.80

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase

C;Keywords: ATP; phosphotransferase; protein kinase

F;136-193/Domain: protein kinase homology (fragment) <KIN>

Query Match 42.7%; Score 979.5; DB 2; Length 421;

Best Local Similarity 49.8%; Pred. No. 4.9e-35;

Matches 206; Conservative 71; Mismatches 116; Indels 21; Gaps 7;

Qy 10 KYEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHP 69

Db 11 RYEVGKFLGQGTFAKVYHARHLKTGDSVAIKVIDKERILKVGTEQIKREISAMRLRHP 70

Qy 70 NIVELYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCHCKG 129

Db 71 NIVELHEVMATKSKIYFVMEHVKGGELEFNK-VSTGKLREDVARKYFQQLVRAVDVCHSRG 129

Qy 130 VYHRDLKPNLLDNTGNLKVSDFGLSALP---QEGVELLRTTCGTPNIVYVAPVLSGQG 185

Db 130 VCHRDLPENLLLDEHGNLKSDFGLSALSDSRRRQDG--LLHTTCGTPAVYVAPVVISRNG 187

Qy 186 YDGSAAADIWSCGVILFVILAGYLPFFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRIL 245

Db 188 YDGFKADVWSCGVILFVLLAGYLPFRDNLNMELYKKIGKAEVKFPNWLAPGAKRLIL 247

Qy 246 DPNPKTRIQIGIKKDPWFLNLYVPIRAREEEVNLDLDIRAVFDGIEGSYVAENVERNDE 305

Db 248 DPNPNTRVSTEKINKSSWFRKG-----LQEVKESVEEETEVDAAEGNASAEKEKKR-- 300

Qy 306 GPLMMNAFEMITLSQGLNLSALFDRRQDFVKKQTRFVSRRRPSSEIITANIEAVANSMGFKS 365

Db 301 -CINLNAFEIISLSTGFDLSGLFEKGE--KEEMRFTSNREASEITEKLAIVEIGKOLKMKV 357

Qy 366 HTRNFKTRLEGSLSSIKAGQLAVVIEIYEVAPSLFMDVVRKAAGETLEYHKFYKK 419

Db 358 RKKEHEWRVK---MSAEATVVEAEVFEIAPSYYHMMVVLKSGGDTAEYKRVKME 407

RESULT 13

G86414

probable protein kinase [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 17-May-2002

C;Accession: G86414

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G86414

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-520 <STO>

A;Cross-references: GB:AE005172; NID:g9502417; PIDN:AAF88116.1; GSPDB:GN00141

C;Genetics:

A;Map position: 1

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo

Query Match 42.7%; Score 978; DB 2; Length 520;

Best Local Similarity 47.0%; Pred. No. 6.8e-35;

Matches 208; Conservative 82; Mismatches 121; Indels 32; Gaps 10;

Qy 8 VGKYEVRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67

Db 71 MGKVELGLLGHGTFKAVYLAQNIKSGDKVAIKVIDKEKIMKSLGVAHIKREISILRRVR 130

Qy 68 HPMIVLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCHC 127

Db 131 HPYVHLFEVMATKSKIYFVMEYVGGELFNITVA-KGRLPETARRYFQQLISSVSFCHG 189

Qy 128 KGVYHRDLKPNLLDNTGNLKVSDFGLSA---LPQEGVELLRTTCGTPNIVYVAPVLSG 183

Db 190 RGVYHRDLKPNLLDNTGNLKVSDFGLSAVALQRLQDG--LCHTFCGTPAYIAPEVLTR 247

Qy 184 QGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHR 243

Db 248 KGYDAAKADVWSCGVILFVLMAGHIPFYDKNIMVMYKKIYKGEFCRPFWSSDLVRLTR 307

Qy 244 ILDPNPKTRIQIGIKKDPWFLNLYVPIR-----AREEEVNLDLDIRAVFDGIEGSY 295

Db 308 LLDTPDTRITIPKIMKNRWFKKGFKHVKFYIEDDKLCREDED---EEEEASSG-RSST 363

Qy 296 VAE-----NVERNDEG-----PLMMNAFEMITLSQGLNLSALFDRRQDFVKKQTRFVSRR 346

Db 364 VESDAEFDVKRMGICSMRPPSSLNADFIIISFSGFDLSGLFEEEG--EGTRFVSGAP 420

Qy 347 PSEIIANIEAVANSMGFKSHTRNFKTRLEGSLSSIKAGQLAVVIEIYEVAPSLFMDVVRKA 406

Db 421 VSKIIISKLEIEIAKIVSFTVRKKWLSRLLEGCRGAKGPLTIAAEIFELTPSLVWVEVKKK 480

Qy 407 AGETLEYHKF-YKKLCSKLENI 428

Db 481 GGDREYEYEFCKNELRPELEKLI 503

RESULT 14  
T14822  
probable serine/threonine protein kinase (EC 2.7.1.1.-) SNFL3 - sorghum  
N;Alternate names: SNF1 protein kinase homolog SNFL3  
C;Species: Sorghum bicolor (sorghum)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
C;Accession: T14822  
R;Annen, F.; Stockhaus, J.  
Bot. Acta 111, 137-142, 1998  
A;Title: SNFL3 a protein kinase homolog of Sorghum bicolor with a high similarity to the  
A;Reference number: Z18217  
A;Accession: T14822  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-461 <AN>  
A;Cross-references: EMBL:Y14274; NID:el301468; PID:el301469  
A;Experimental source: cultivar TX 430  
C;Genetics:  
A;Note: SNFL3  
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C;Keywords: phosphotransferase; protein kinase  
F;10-266/Domain: protein kinase homolog <KIN>  
  
Query Match 42.2%; Score 968.5; DB 2; Length 461;  
Best Local Similarity 46.4%; Pred. No. 1.5e-34;  
Matches 202; Conservative 85; Mismatches 121; Indels 27; Gaps 9;  
  
Qy 10 KYEVGRITGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHP 69  
Db 11 RYEIGRQLGQGNFAKVKFARNLTDGQSAIKMIDKDKITRVLIVQIKREISIMRLVKHP 70  
  
Qy 70 NIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCHCKG 129  
Db 71 NVLQLFEVMASKSIYIVLEFVTGGELFNKI-SKGKFSQEDVARRYFQQLISAVDYCHSRG 129  
  
Qy 130 VYHRDLKPENLLDNGNLKVSDFGLSALPQ---EGVELLRTTCGTPNYVAPEVLSGQG 185  
Db 130 VYHRDLKPENLLDENENLKVSDFGLSALAESKRHDG--LLHTTCGTPAYVAPEVLSRRG 187  
  
Qy 186 YDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHRIL 245  
Db 188 YDGTKADIWSCGVILFVILAGYLPFHPDPLNLIEMVYKISKAEYRCPRFSGELKDLLFRML 247  
  
Qy 246 DENPKTRIQIOGIKKDPWFRNLNVPIRAREEEVNLDI---RAVFDGIEGSYVAENVE 301  
Db 248 DEDPSTRISIRIKRSTWYKPIEVNSAKIKHDTTRDKVCNGEATTSNIECSNSEET-- 305  
  
Qy 302 RNDEGLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 361  
Db 306 RGSSSLPLNINAFDIIISLTGFDLSNLFEEK--YGRAREERTTRQAGTIFAKLNELAE 363  
  
Qy 362 GPKSHTRNFKTRLEGSLSIKA-----GQLAVVIEIYEVAPSLEFMDVVRKAAGETLEYHK 415  
Db 364 KLK-----IKKENGVLKASPKEGKIGFLELDABIFELAPSFLLVLEKKTNGDTIEYQR 418  
  
Qy 416 FYK-KLCSKLENIW 429  
Db 419 LVKEIRPALKDMVW 433  
  
RESULT 15  
T02306  
probable protein kinase [imported] - Arabidopsis thaliana  
N;Alternate names: hypothetical protein F13P17.2  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 16-Feb-2001  
C;Accession: T02306; D84753  
R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, July 1998  
A;Description: Arabidopsis thaliana chromosome II BAC F13P17 genomic sequence.  
A;Reference number: Z14657  
A;Accession: T02306

A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-502 <ROU>  
A;Cross-references: EMBL:AC004481; NID:g3337347; PIDN:AAC27394.1; PID:g3337349  
A;Experimental source: cultivar Columbia  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: D84753  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-502 <STO>  
A;Cross-references: GB:AE002093; NID:g3337349; PIDN:AAC27394.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g34180; F13P17.2  
A;Map position: 2  
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
F;55-311/Domain: protein kinase homolog <KIN>  
  
Query Match 40.8%; Score 935; DB 2; Length 502;  
Best Local Similarity 44.7%; Pred. No. 4.3e-33;  
Matches 199; Conservative 91; Mismatches 123; Indels 32; Gaps 11;  
  
Qy 10 KYEVGRITGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHP 69  
Db 56 KYEIGKLLGHGSAKVKYLARNIHSGEDVAIKVIDKEIKVSGLAGHIKREISILRRVRHP 115  
  
Qy 70 NIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCHCKG 129  
Db 116 YIVHLLVEMATKTIYIVMEYVVRGGELYNTVA-RGLRREGTARRYFQQLISSVAFCHSRG 174  
  
Qy 130 VYHRDLKPENLLDNGNLKVSDFGLSA---LPQEGVELLRTTCGTPNYVAPEVLSGQG 185  
Db 175 VYHRDLKLENLLDDKGNVKVSDFGLSVVSEQLKQEGI--CQTFCGTPAYLAPEVLTTRKG 232  
  
Qy 186 YDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHRIL 245  
Db 233 YEGAKADIWSCGVILFVILAGYLPFDDDKNILVMYTKIYKQKFCPKWFSPELARLVTRML 292  
  
Qy 246 DENPKTRIQIOGIKKDPWFRNLNVPIR-----AREEEVNLDIIRAVFDGIEGSYVA 297  
Db 293 DTNPDTTRITPEIMKHWRWFKGFKHVKFYIENDKLCREDDDDDDSSSLSSG-RSSTAS 351  
  
Qy 298 E-----NVERNDE--GPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEI 350  
Db 352 EGDAEFDIKRVDSMPREASLNADFILSFS---DLGLFEEG---GQGARFVSAAPMTKI 404  
  
Qy 351 IANIEAVANSMGFKSHTRNFKTRLEGSLSIKACQLAVVIEIYEVAPSLEFMDVVRKAAGET 410  
Db 405 ISKLEEIAKEVKFVVRKKDWSVRLEGCGREGAKGKPLTIRVEIFELTPSLVVVEVKKGGNI 464  
  
Qy 411 LEYHKF-YKKLCSKLENIW-WRATE 433  
Db 465 EYEFEFCNKELRPQLEKLMHYQADE 489  
  
Search completed: October 20, 2003, 16:52:20  
Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:52:02 ; Search time 69 Seconds  
(without alignments)  
1059.522 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTKMRRVGKVEVGRITIGEG.....IIWRATEGIPKSEILRTITF 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 609560 seqs, 163917102 residues  
Total number of hits satisfying chosen parameters: 609560

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2293	100.0	446	9	US-09-824-735-2
2	1545	67.4	450	12	US-10-292-408-9
3	669	29.2	532	9	US-09-824-735-4
4	663	28.9	633	9	US-09-824-735-3
5	663	28.9	633	10	US-09-801-368-338
6	608.5	26.5	1203	9	US-09-799-875-5
7	602	26.3	778	12	US-10-354-358-92
8	602	26.3	778	12	US-10-116-326-2
9	593.5	25.9	651	9	US-09-870-937-10
10	593.5	25.9	651	10	US-09-974-298-112
11	593.5	25.9	651	12	US-10-354-358-8
12	591	25.8	793	15	US-10-195-101-32
13	589.5	25.7	1518	10	US-09-801-368-152
14	588	25.6	795	10	US-09-919-585-12
15	588	25.6	795	12	US-10-161-565-25

16	588	25.6	795	15	US-10-142-356-9	Sequence 9, Appli
17	586.5	25.6	798	11	US-09-823-187-95	Sequence 95, Appl
18	586.5	25.6	832	10	US-09-919-585-21	Sequence 21, Appl
19	586	25.6	752	10	US-09-835-081-2	Sequence 2, Appli
20	585	25.5	744	10	US-09-919-585-3	Sequence 3, Appli
21	584	25.5	729	12	US-10-161-565-26	Sequence 26, Appl
22	584	25.5	729	15	US-10-142-356-11	Sequence 11, Appl
23	584	25.5	729	15	US-10-195-101-33	Sequence 33, Appl
24	583	25.4	744	10	US-09-835-081-4	Sequence 4, Appli
25	579.5	25.3	779	8	US-08-817-832B-31	Sequence 31, Appl
26	578	25.2	688	12	US-10-161-565-28	Sequence 28, Appl
27	578	25.2	688	12	US-10-161-565-29	Sequence 29, Appl
28	578	25.2	713	12	US-10-161-565-27	Sequence 27, Appl
29	576.5	25.1	776	11	US-09-823-187-92	Sequence 92, Appl
30	576.5	25.1	776	11	US-09-823-187-93	Sequence 93, Appl
31	576.5	25.1	776	15	US-10-195-101-34	Sequence 34, Appl
32	575	25.1	691	10	US-09-919-585-6	Sequence 6, Appli
33	575	25.1	722	8	US-08-817-832B-32	Sequence 32, Appl
34	575	25.1	724	10	US-09-919-585-9	Sequence 9, Appli
35	575	25.1	745	12	US-10-161-565-24	Sequence 24, Appl
36	575	25.1	745	15	US-10-195-101-36	Sequence 36, Appl
37	574	25.0	668	14	US-10-054-579-2	Sequence 2, Appli
38	574	25.0	668	15	US-10-195-072-2	Sequence 2, Appli
39	574	25.0	668	15	US-10-195-071-2	Sequence 2, Appli
40	574	25.0	674	15	US-10-283-247-2	Sequence 2, Appli
41	573	25.0	783	9	US-09-815-915-2	Sequence 2, Appli
42	573	25.0	783	12	US-10-393-316-2	Sequence 2, Appli
43	570	24.9	674	10	US-09-842-582-2	Sequence 2, Appli
44	570	24.9	674	15	US-10-283-247-7	Sequence 7, Appli
45	570	24.9	674	15	US-10-283-247-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-09-824-735-2  
; Sequence 2, Application US/09824735  
; Patent No. US20020095032A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHU, JIAN-KANG  
; APPLICANT: LIU, JIPING  
; APPLICANT: ISHITANI, MANABU  
; APPLICANT: HALFTER, URSULA  
; APPLICANT: KIM, CHEOL-SOO  
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS  
; FILE REFERENCE: 205645US20  
; CURRENT APPLICATION NUMBER: US/09/824,735  
; CURRENT FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: US 60/824,735  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-824-735-2

Query Match	100.0%;	Score 2293;	DB 9;	Length 446;
Best Local Similarity	100.0%;	Pred. No. 2.4e-193;		
Matches	446;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	1	MTKMRVRVGKVEVGRITIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREI	60	
Db	1	MTKMRVRVGKVEVGRITIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREI	60	
Qy	61	SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVD	120	
Db	61	SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVD	120	
Qy	121	AVAHCHCKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYVAPEV	180	



Db 121 AVAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRITTCGTPNYVAPEV 180

QY 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240

Db 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240

QY 241 IHRILDPNPKTRIQIQIKKDPWFRLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300

Db 241 IHRILDPNPKTRIQIQIKKDPWFRLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300

QY 301 ERNDEGPLMWNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360

Db 301 ERNDEGPLMWNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360

QY 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420

Db 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKL 420

QY 421 CSKLENIWRATEGIPKSEILRTITF 446

Db 421 CSKLENIWRATEGIPKSEILRTITF 446

RESULT 2

US-10-292-408-9

; Sequence 9, Application US/10292408

; Publication No. US20030182692A1

; GENERAL INFORMATION:

; APPLICANT: VAN THIELEN, NOCHA

; APPLICANT: DA COSTA E SILVA, OSWALDO

; APPLICANT: CHEN, RUOYING

; TITLE OF INVENTION: PROTEIN KINASE STRESS-RELATED POLYPEPTIDES AND METHODS

; TITLE OF INVENTION: OF USE IN PLANTS

; FILE REFERENCE: 16313-0178

; CURRENT APPLICATION NUMBER: US/10/292,408

; CURRENT FILING DATE: 2002-11-12

; PRIOR APPLICATION NUMBER: 60/346,096

; PRIOR FILING DATE: 2001-11-09

; NUMBER OF SEQ ID NOS: 51

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 9

; LENGTH: 450

; TYPE: PRT

; ORGANISM: Physcomitrella patens

us-10-292-408-9

Query Match 67.4%; Score 1545; DB 12; Length 450;

Best Local Similarity 66.0%; Pred. No. 1.5e-127;

Matches 293; Conservative 70; Mismatches 75; Indels 6; Gaps 4;

QY 5 MRRVGKYEVRTIGEGTFAKVKFARNRDTGDNVAIKIMAKSTILKNRMVDQIKREISIMK 64

Db 3 MRKVGKYEVRTIGEGTFAKVKFAQNTETGESVAMKVLDRQTLKHKMVEQIRREISIMK 62

QY 65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKXYFQQLVDAVAH 124

Db 63 LVRHPNVVRLHEVLASRCKIYIILEFVTGGELFDKI VHQGRLNENDSRKYFQQLMDGVDY 122

QY 125 CHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVE--LLRITTCGTPNYVAPEVLS 182

Db 123 CHSKGVSHRDLKPENLLDLSLDNLKISDFGLSALPQQVREPLLHTTCGTPNYVAPEVLN 182

QY 183 GQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKELIH 242

Db 183 DKGYDGAVADIWSCGVILFVLMAGFLPFDEADLNTLYSKIREADFTCPPWFSSGAKTLIT 242

QY 243 RILDPNPKTRIQIQIKKDPWFRLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 302

Db 243 NILDPNPLTRIMRGIRDDEWFKNYVPVRMYDDEDINLDDVETAFDDSKQEVKREQREV 302

QY 303 NDEGPLMWNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 362

Db 303 KDVGPSLMNFAFELISLSQGLNLSALFDRRQDHVKRQTRFTSKKPARDIINRMETAASKSMG 362

QY 363 FKSHTRNFKTRLEGLSSIKAGQ-LAVVIEIYEVAPSLFMVDVRKAAGETLEYHKFYKKLC 421

Db 363 FGVGTRNFKTRLEAASECRISQHLAVAIEVYEVAPSLFMIEVRKAAGDTLEYHKFYKSFC 422

QY 422 SKLENIWRATEGIPKSEILRTIT 445

Db 423 TRLKDIIW--TTAVDKDEV-KTLT 443

RESULT 3

US-09-824-735-4

; Sequence 4, Application US/09824735

; Patent No. US20020095032A1

; GENERAL INFORMATION:

; APPLICANT: ZHU, JIAN-KANG

; APPLICANT: LIU, JIPING

; APPLICANT: ISHITANI, MANABU

; APPLICANT: HALFTER, URSULA

; APPLICANT: KIM, CHEOL-SOO

; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS

; FILE REFERENCE: 205645US20

; CURRENT APPLICATION NUMBER: US/09/824,735

; CURRENT FILING DATE: 2001-08-17

; PRIOR APPLICATION NUMBER: US 60/824,735

; PRIOR FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 552

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-824-735-4

Query Match 29.2%; Score 669; DB 9; Length 552;

Best Local Similarity 36.9%; Pred. No. 2.4e-50;

Matches 151; Conservative 77; Mismatches 147; Indels 34; Gaps 7;

QY 1 MTKKMR-----RVGKYEVRTIGEGTFAKVKFARNRDTGDNVAIKIMAKSTILKNRMVDQ 55

Db 1 MAEKQKHGRVKIGHYVLGTLGVTGFKVKIGEHQLTGHKVAVKILNRQKIRSLDVVGK 60

QY 56 IKREISIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKXYF 115

Db 61 IKREIQNLKLFRRPHIILKYIVISTPTDFFMWMEYVSGGELFDYICKHGRVEEMEARRLF 120

QY 116 QQLVDVAVAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRITTCGTPNY 175

Db 121 QQILSAVDYCHRRHVVHRDLKPENVLDDAHMNAKIADFGLSNMMSDG-EFLRTSCSPNY 179

QY 176 VAPEVLSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSA 235

Db 180 AAPEVISGRLYAGPEVDIWSCGVILYALLCGTLFDDDEHVPTLFFKKIRGGVFYIPEYLN 239

QY 236 EVKFLIHRILDPNPKTRIQIQIKKDPWFRLNYPVIRAREEE--EVNLDLDIRAVFDGIE- 292

Db 240 SVATLLMHMLQVDPPLKRATIKDIREHEWFKQGLPSYLPFDPEDPSYDANVIDDEAVKEVCEK 299

QY 293 ----GSYVAENVERNDEGPLMWNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPS 348

Db 300 FECTESEVMNSLYSGDPQDLAVAYHLI-----IDNRRIMNQASEFYLAASSPPS 348

QY 349 EIIANIEAVANSMGFKSHTRNF-----KTR--LEGLSSIKAGQLAV 387

Db 349 GSFMDDSAMHIPPLGLKPHPERMPPLIADSPKARCPDLALNTTKPKSLAV 397

RESULT 4

US-09-824-735-3

; Sequence 3, Application US/09824735

; Patent No. US20020095032A1

; GENERAL INFORMATION:

; APPLICANT: ZHU, JIAN-KANG

```
; APPLICANT: LIU, JIPING
; APPLICANT: ISHITANI, MANABU
; APPLICANT: HALFTER, URSULA
; APPLICANT: KIM, CHEOL-SOO
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS
; FILE REFERENCE: 205645US20
; CURRENT APPLICATION NUMBER: US/09/824,735
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/824,735
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-824-735-3

Query Match      28.9%; Score 663; DB 9; Length 633;
Best Local Similarity 39.4%; Pred. No. 9.5e-50;
Matches 135; Conservative 80; Mismatches 98; Indels 30; Gaps 6;

Qy      8 VGKYEVRTIGEGTFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVR 67
Db      52 IGNYQIVKTLGEGSFGKVKLAYHTTTGQKVALKIINKVLAKSDMQGRIEREISYLRLLR 111
Qy      68 HENIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESESRKYFQQLVDVAHCHC 127
Db      112 HPHIILYDVIKSKDEIIMVIEY-AGNELFDYIVQDKMSEQEARRFFQIISAVEYCHR 170
Qy      128 KGVYHRDLKPENLLDNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYD 187
Db      171 HKIVHRDLKPENLLDDEHLNVKIADFGLSNIMTDG-NFLKTSCGSPNYAAPEVISGKLYA 229
Qy      188 GSAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDP 247
Db      230 GPEVDVWSCGVILYVMLCRRLPFDDESIPVLFKNISNGVYTLPKFLSPGAAGLIKRMILV 289
Qy      248 NPKTRIQIQIKKDPWFLN---YV---PIRAREEE-----VNLD 282
Db      290 NPLNRISIHEIMQDDWFKVDLPPEYLLPPDLKPHPEEENENNDKKGSSPDNDDEIDNLV 349
Qy      283 DIRAVFDGIEGSYVAENVERNDEGPL---MMNAFEMITLSQGL 322
Db      350 NILSSTMGYEKDEIYESLESSEDTPAFNEIRDAYMLIKENKSL 392

RESULT 5
US-09-801-368-338
; Sequence 338, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250Alman, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
```

```
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 338
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-338

Query Match      28.9%; Score 663; DB 10; Length 633;
Best Local Similarity 39.4%; Pred. No. 9.5e-50;
Matches 135; Conservative 80; Mismatches 98; Indels 30; Gaps 6;

Qy      8 VGKYEVRTIGEGTFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVR 67
Db      52 IGNYQIVKTLGEGSFGKVKLAYHTTTGQKVALKIINKVLAKSDMQGRIEREISYLRLLR 111
Qy      68 HENIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESESRKYFQQLVDVAHCHC 127
Db      112 HPHIILYDVIKSKDEIIMVIEY-AGNELFDYIVQDKMSEQEARRFFQIISAVEYCHR 170
Qy      128 KGVYHRDLKPENLLDNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYD 187
Db      171 HKIVHRDLKPENLLDDEHLNVKIADFGLSNIMTDG-NFLKTSCGSPNYAAPEVISGKLYA 229
Qy      188 GSAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDP 247
Db      230 GPEVDVWSCGVILYVMLCRRLPFDDESIPVLFKNISNGVYTLPKFLSPGAAGLIKRMILV 289
Qy      248 NPKTRIQIQIKKDPWFLN---YV---PIRAREEE-----VNLD 282
Db      290 NPLNRISIHEIMQDDWFKVDLPPEYLLPPDLKPHPEEENENNDKKGSSPDNDDEIDNLV 349
Qy      283 DIRAVFDGIEGSYVAENVERNDEGPL---MMNAFEMITLSQGL 322
Db      350 NILSSTMGYEKDEIYESLESSEDTPAFNEIRDAYMLIKENKSL 392

RESULT 6
US-09-799-875-5
; Sequence 5, Application US/09799875
; Patent No. US20020034780A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Williamson, Mark
; TITLE OF INVENTION: No. US20020034780A1el Human Protein Kinases and Uses
; TITLE OF INVENTION: Therefor
; FILE REFERENCE: 35800/209996
; CURRENT APPLICATION NUMBER: US/09/799,875
; CURRENT FILING DATE: 2001-03-06
; PRIOR APPLICATION NUMBER: 60/182,059
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/659,287
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1203
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-799-875-5

Query Match      26.5%; Score 608.5; DB 9; Length 1203;
Best Local Similarity 37.9%; Pred. No. 1.4e-44;
Matches 146; Conservative 69; Mismatches 121; Indels 49; Gaps 12;

Qy      7 RVGKYEVRTIGEGTFAKVKFARNTDTGDNVAIKMAKSTILKNRMVDQIKREISIMKIV 66
Db      4 RIGYEIDRTIGKGNFAVVKRATHLVTKAKVAIKIDK-TQLDEENLKKIFREVQIMKML 62
Qy      67 RHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLSESESRKYFQQLVDVAHCH 126
Db      63 CHPHIIRLYQVMETERMIYLVTEYASGGEIFDLVHAGRMAEKEARRKFKQIVTAVYFCH 122
```





RESULT 9  
US-09-870-937-10  
; Sequence 10, Application US/09870937  
; Patent No. US20020049180A1  
; GENERAL INFORMATION:  
; APPLICANT: Wu, Bin  
; APPLICANT: Seeley, Todd  
; APPLICANT: Williams, Lewis T.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING NEOPLASTIC DISEASE USING  
; TITLE OF INVENTION: CHEMOTHERAPY AND RADIATION SENSITIZERS  
; FILE REFERENCE: 200130.514/PP-01623.002  
; CURRENT APPLICATION NUMBER: US/09/870,937  
; CURRENT FILING DATE: 2001-05-30  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 10  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-870-937-10

Query Match 25.9%; Score 593.5; DB 9; Length 651;  
Best Local Similarity 43.8%; Pred. No. 1.3e-43;  
Matches 121; Conservative 58; Mismatches 92; Indels 5; Gaps 4;  
  
QY 11 YEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHPN 70  
Db 11 YELHETIGTGFAKVKLACHILTGEMVAIKINDKNTLGS--LPRIKTEIEALKNLRHQH 68  
  
QY 71 IVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYPQQLVDAVAHCHCKGV 130  
Db 69 ICQLYHVLEETANKIFMVLEYCPGGELFDYIISQDRLSEETRVVFRQIVSAVAVVHSQGY 128  
  
QY 131 YHRDLKPENLLDNTGNLKVSDFGLSALPQEGVEL-LRTTCGTPNVVAPEVLSGQGYDGS 189  
Db 129 AHRDLKPENLLDFEYHKLLIDFGLCAKPKGNKDYLHLQTCGSLAYAAPELIQKSYLGS 188  
  
QY 190 AADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPNP 249  
Db 189 EADVWSMGILLYVLMCGFLPFDDDDNMALYKKIMRGKYDVPKWLSPSSILLLOQMLQVDP 248  
  
QY 250 KTRIQUIQGIKKDPWFRNLNY-VPIRAREEEEE-VNLDD 283  
Db 249 KKRISMKNLLNHPWIMQDYNYPVWQSKNPFHLLDD 284

RESULT 10  
US-09-974-298-112  
; Sequence 112, Application US/09974298  
; Patent No. US20020156263A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hwei-Mei  
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
; FILE REFERENCE: PA-0037 P  
; CURRENT APPLICATION NUMBER: US/09/974,298  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 194  
; SOFTWARE: PERL Program  
; SEQ ID NO 112  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20020156263A1 256716CD1  
US-09-974-298-112

Query Match 25.9%; Score 593.5; DB 10; Length 651;  
Best Local Similarity 43.8%; Pred. No. 1.3e-43;  
Matches 121; Conservative 58; Mismatches 92; Indels 5; Gaps 4;

QY 11 YEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRHPN 70  
Db 11 YELHETIGTGFAKVKLACHILTGEMVAIKINDKNTLGS--LPRIKTEIEALKNLRHQH 68  
  
QY 71 IVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYPQQLVDAVAHCHCKGV 130  
Db 69 ICQLYHVLEETANKIFMVLEYCPGGELFDYIISQDRLSEETRVVFRQIVSAVAVVHSQGY 128  
  
QY 131 YHRDLKPENLLDNTGNLKVSDFGLSALPQEGVEL-LRTTCGTPNVVAPEVLSGQGYDGS 189  
Db 129 AHRDLKPENLLDFEYHKLLIDFGLCAKPKGNKDYLHLQTCGSLAYAAPELIQKSYLGS 188  
  
QY 190 AADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPNP 249  
Db 189 EADVWSMGILLYVLMCGFLPFDDDDNMALYKKIMRGKYDVPKWLSPSSILLLOQMLQVDP 248  
  
QY 250 KTRIQUIQGIKKDPWFRNLNY-VPIRAREEEEE-VNLDD 283  
Db 249 KKRISMKNLLNHPWIMQDYNYPVWQSKNPFHLLDD 284

RESULT 11  
US-10-354-358-8  
; Sequence 8, Application US/10354358  
; Publication No. US20030157082A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc  
; APPLICANT: Hunter, John Joseph  
; APPLICANT: MacBeth, Kyle J.  
; APPLICANT: Tsai, Fong-Ying  
; APPLICANT: Lesoon, Andrea  
; APPLICANT: Lightcap, Eric S.  
; APPLICANT: Williamson, Mark  
; APPLICANT: Rudolph-Owen, Laura A.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
; TITLE OF INVENTION: CANCER USING 140, 1470, 1686, 2089, 2427, 3702, 5891, 6428,  
; TITLE OF INVENTION: 7181, 7660, 25641, 69583, 49863, 8897, 1682, 17667, 9235,  
; TITLE OF INVENTION: 3703, 14171, 10359, 1660, 1450, 18894, 2088, 32427, 2160,  
; TITLE OF INVENTION: 9252, 9389, 1642, 85269, 10297, 1584, 9525, 14124, 4469,  
; TITLE OF INVENTION: 8990, 2100, 9288, 64698, 10480, 20893, 33230, 1586, 9943,  
; TITLE OF INVENTION: 16334, 68862, 9011, 14031, 6178, 21225, 1420, 32236, 2099,  
; FILE REFERENCE: MPI02-020P1RNMNMIM  
; CURRENT APPLICATION NUMBER: US/10/354,358  
; CURRENT FILING DATE: 2003-01-30  
; PRIOR APPLICATION NUMBER: US 60/353,600  
; PRIOR FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: US 60/364,517  
; PRIOR FILING DATE: 2002-03-15  
; PRIOR APPLICATION NUMBER: US 60/371,075  
; PRIOR FILING DATE: 2002-04-09  
; PRIOR APPLICATION NUMBER: US 60/371,507  
; PRIOR FILING DATE: 2002-04-10  
; PRIOR APPLICATION NUMBER: US 60/372,984  
; PRIOR FILING DATE: 2002-04-16  
; PRIOR APPLICATION NUMBER: US 60/374,194  
; PRIOR FILING DATE: 2002-04-19  
; PRIOR APPLICATION NUMBER: US 60/382,995  
; PRIOR FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/385,023  
; PRIOR FILING DATE: 2002-05-31  
; PRIOR APPLICATION NUMBER: US 60/388,853  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: US 60/389,395  
; PRIOR FILING DATE: 2002-06-17  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 122  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 651  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-354-358-8

Query Match 25.9%; Score 593.5; DB 12; Length 651;  
Best Local Similarity 43.8%; Pred. No. 1.3e-43;  
Matches 121; Conservative 58; Mismatches 92; Indels 5; Gaps 4;

QY 11 YEVGRTIGETFAKVKFARNTDGTGDNVAIKMAKSTILKNRMVDQIKREISIMKIVRHPN 70  
Db 11 YELHETIGTGGAFAKVKLACHILTGENVAIKINDKNTLGSDDLPRIKTEIEALKNLRHQH 68  
QY 71 IVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDAVAHCHCKGV 130  
Db 69 ICQLYHVLETAANKIFMVLEPCPGGELFDYIISQDRLSEETRVVFRQIVSAVAVHSQGY 128  
QY 131 YHRDLKPENLLDNTGNLKVSDFGLSALPQEGVEL-LRTTCGTPNYVAPEVLSGQGYDGS 189  
Db 129 AHRDLKPENLLDDEYHKLIDFGLCAKPKGNKDYHLQTCGSLAYAAPELIQKSYLGS 188  
QY 190 AADIWSCGVILFVILAGYLPFSETDPLGLYRKINAEEFSCPPWFSAEVKFLIHRILDPNP 249  
Db 189 EADVWSMGILLYVLMCGFLPFDDDNVMALYKIMRGKYDVPKWLSPSSILLQOMLQVDP 248  
QY 250 KTRIQIGIKKDPWFRINY-VPIRAREEEE-VNLDD 283  
Db 249 KKRISMKNLNLNHPWIMQDYNYPVEWQSKNPFILHDD 284

RESULT 12

US-10-195-101-32  
; Sequence 32, Application US/10195101  
; Publication No. US20030087317A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Molteni, Angela  
; APPLICANT: Magnaghi, Paola  
; APPLICANT: Bosotti, Roberta  
; APPLICANT: Scacheri, Emanuela  
; APPLICANT: Isacchi, Antonella  
; APPLICANT: Hodgson, David M.  
; TITLE OF INVENTION: HUMAN NIM1 KINASE  
; FILE REFERENCE: PC-0009-1 CIP  
; CURRENT APPLICATION NUMBER: US/10/195,101  
; CURRENT FILING DATE: 2002-07-11  
; PRIOR APPLICATION NUMBER: 09/523,849  
; PRIOR FILING DATE: 2000-03-13  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PERL Program  
; SEQ ID NO 32  
; LENGTH: 793  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; NAME/KEY: misc.feature  
; OTHER INFORMATION: GenBank Accession No. US20030087317A1 g2052189  
; DATABASE ENTRY DATE: 25 April 1997

US-10-195-101-32

Query Match 25.8%; Score 591; DB 15; Length 793;  
Best Local Similarity 35.1%; Pred. No. 2.8e-43;  
Matches 135; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

QY 2 TKKMRVGVKYEVRTIGETFAKVKFARNTDGTGDNVAIKMAKSTILKNRMVDQIKREIS 61  
Db 51 TDEQPHICNRYLRQKTIGKNFAKVKLARHVLTRGAVAVKIIDK-TQLNPTSLQKLFREVR 109  
QY 62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDA 121  
Db 110 IMKILNHPNIVKLFEVETETKTLVWMEYASGGEVFDYLVAGHGRMKEARAKFRQIVSA 169  
QY 122 VAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVL 181  
Db 170 VQYCHQKIVHRDLKAENLLLDADNMNIADFGFSNEFTVG-NKLDFTCGSPPYAAPELF 228

QY 182 SQQYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAEEFSCPPWFSAEVKFLI 241  
Db 229 QGKKYDGPEDVWSLGLVILYTLVSGSLPFDGQNLKELRERVLGRKYRVFFYMTDCENLL 288  
QY 242 HRILDPNPKTRIQIGIKKDPWFRINY-VPIRAREEEEVNLLDIRAV----- 287  
Db 289 KKLVLNLPKRGSLAQIMKDRWMNVGHEEEELKPYSEPELDLNDAKRIDIMVTMGFARDE 348  
QY 288 -----FDGIEGSYVA-----ENVERNDEGPLMMNAFEMITLSQGLNLSAL--- 327  
Db 349 INDALVSQKYDEVMATYILLGRKPPFEFGESLSSG-----NLCQSRPSSDLNNTLQSP 404  
QY 328 ----FDRQDFVKRQTRFVSRREPS 348  
Db 405 AHLKVQRSISANQKORRFSDHAGPS 429

RESULT 13

US-09-801-368-152  
; Sequence 152, Application US/09801368  
; Patent No. US20020128250A1  
; GENERAL INFORMATION:  
; APPLICANT: Busby, Robert  
; APPLICANT: Cali, Brian  
; APPLICANT: Hecht, Peter  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin  
; APPLICANT: Maxon, Mary  
; APPLICANT: Milne, Todd  
; APPLICANT: No. US20020128250A1man, Thea  
; APPLICANT: Royer, John  
; APPLICANT: Salama, Sofie  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeff  
; APPLICANT: Summers, Eric  
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi  
; FILE REFERENCE: 109272.147  
; CURRENT APPLICATION NUMBER: US/09/801,368  
; CURRENT FILING DATE: 2001-03-07  
; PRIOR APPLICATION NUMBER: US 09/487,558  
; PRIOR FILING DATE: 2000-01-19  
; PRIOR APPLICATION NUMBER: US 60/160,587  
; PRIOR FILING DATE: 1999-10-20  
; NUMBER OF SEQ ID NOS: 440  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 152  
; LENGTH: 1518  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
; US-09-801-368-152

Query Match 25.7%; Score 589.5; DB 10; Length 1518;  
Best Local Similarity 30.7%; Pred. No. 8.9e-43;  
Matches 138; Conservative 90; Mismatches 135; Indels 87; Gaps 10;

QY 2 TKKMR-----VGKYEVRTIGETFAKVKFARNTDGTGDNVAIKMAKSTIL----- 48  
Db 67 TKSSKRKSRDVTGPPWKLGTGKSSGRVRLAKNMETQLAAIKIVPKKAFVHCSNNGT 126  
QY 49 -----KNRMVDQ-----IKREISIMKIVRHPNIVRLYEVLASP 81  
Db 127 VPNSYSSSMVTSNVSSPSIASREHSNHSQTNPYGIEREIVIMKLSHTNVMALFEVWENK 186  
QY 82 SKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDAVAHCHCKGVYHRDLKPENLL 141  
Db 187 SELYLVLVEYVDGGELEFDYLVSKGKLPEREAIHYFKQIVEGVSYCHSFNICHRLDKPENLL 246  
QY 142 LD-TNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYDGSADIWSCGVIL 200  
Db 247 LDKNRRRIKIADFGMAALELPN-KLLKLTSCGSPHYASPEIVMGRPHYGGPSDWSGIVL 305  
QY 201 FVILAGYLPFSETDPLGLYRKINAEEFSCPPWFSAEVKFLIHRILDPNPKTRIQIGIKK 260

Db 306 FALLTGHLPFNDDNIKLLLVQSGKYQMPNSLSSEARDLISKILVIDPEKRITTBQILK 365  
QY 261 DPWFR-----LNVVPIRABEE-----EEVNLDI 284  
Db 366 HPLIKKYDDLVPVKVLRKMRKDNMARGKNSDLHLNNVSPSIVTLHSGEIDESILRSL 425  
QY 285 RAVFDGIEGSVAENVERN--DEGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFV 342  
Db 426 QILWHGVSRELITAKLLQKPMSEKLFYSLLLQYKQRHSISLSSSEN-----KKSATES 480  
QY 343 SRREPSEIIANIEAVANSMGFKSHTRNFKT 372  
Db 481 SVNEPR--IEYASKTANNTGLRSENNDVKT 508

RESULT 14  
US-09-919-585-12  
; Sequence 12, Application US/09919585  
; Patent No. US20020115167A1  
; GENERAL INFORMATION:  
; APPLICANT: Sun, Tian-Qiang  
; APPLICANT: Feng, Jia-Jia  
; APPLICANT: Reinhard, Christoph  
; APPLICANT: Fantl, Wendy J.  
; APPLICANT: Williams, Lewis T.  
; TITLE OF INVENTION: ISOLATION OF DROSOPHILA AND HUMAN POLYNUCLEOTIDES ENCODING PAR-1  
; TITLE OF INVENTION: POLYPEPTIDES ENCODED BY THE POLYNUCLEOTIDES AND POLYPEPTIDES  
; FILE REFERENCE: PP-016093.002/200130.525  
; CURRENT APPLICATION NUMBER: US/09/919,585  
; CURRENT FILING DATE: 2001-07-30  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 795  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-585-12

Query Match 25.6%; Score 588; DB 10; Length 795;  
Best Local Similarity 34.9%; Pred. No. 5.1e-43;  
Matches 135; Conservative 73; Mismatches 125; Indels 54; Gaps 8;  
QY 2 TKMRRVGVKYEVRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREIS 61  
Db 51 TDEQPHIGNYRLQKTIGKGNFAKVLARHVLGTREAVAVKIIDK-TQLNPTSLQKLFREVR 109  
QY 62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDA 121  
Db 110 IMKILNHPNIVKLFEVETETKTLVMEYASGGEVFDYLVAHGRMKEKEARAKFRQIVSA 169  
QY 122 VAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVL 181  
Db 170 VQYCHQKYIVHRDLKAENLLDGDNMNIKIADFGFSNEFTVG-NKLDTFGSPPPYAAPELF 228  
QY 182 SGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLI 241  
Db 229 QGKKYDGPEDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLL 288  
QY 242 HRILDPNPKTRIQQIGIKKDPWFRNLNVPIRAREEEVNLDIIRAVFDGIEGSYVAENVE 301  
Db 289 KLLVLNPIKRGSLQIMKDRWMNVGH-----EEEEELK-----PYTEPPDP 329  
QY 302 RNDEGPLMMNAFEMITLS---QGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVA 358  
Db 330 FNDTKRIDI----MVTMGFARDEIN-DALINQKYDEVMATYILLGRKPP----- 373  
QY 359 NSMGFKSHTRNFKTRLEGLSSIKAGQL 385  
Db 374 -----EFEGGESLSSGNL 386

RESULT 15

US-10-161-565-25  
; Sequence 25, Application US/10161565  
; Publication No. US20030165809A1  
; GENERAL INFORMATION:  
; APPLICANT: EXELIXIS, INC.  
; TITLE OF INVENTION: MARKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE  
; FILE REFERENCE: EX02-088C  
; CURRENT APPLICATION NUMBER: US/10/161,565  
; CURRENT FILING DATE: 2002-06-03  
; PRIOR APPLICATION NUMBER: US 60/296,076  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/328,605  
; PRIOR FILING DATE: 2001-10-10  
; PRIOR APPLICATION NUMBER: US 60/357,253  
; PRIOR FILING DATE: 2002-02-15  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 795  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-161-565-25

Query Match 25.6%; Score 588; DB 12; Length 795;  
Best Local Similarity 34.9%; Pred. No. 5.1e-43;  
Matches 135; Conservative 73; Mismatches 125; Indels 54; Gaps 8;  
QY 2 TKMRRVGVKYEVRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREIS 61  
Db 51 TDEQPHIGNYRLQKTIGKGNFAKVLARHVLGTREAVAVKIIDK-TQLNPTSLQKLFREVR 109  
QY 62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDA 121  
Db 110 IMKILNHPNIVKLFEVETETKTLVMEYASGGEVFDYLVAHGRMKEKEARAKFRQIVSA 169  
QY 122 VAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVL 181  
Db 170 VQYCHQKYIVHRDLKAENLLDGDNMNIKIADFGFSNEFTVG-NKLDTFGSPPPYAAPELF 228  
QY 182 SGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLI 241  
Db 229 QGKKYDGPEDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLL 288  
QY 242 HRILDPNPKTRIQQIGIKKDPWFRNLNVPIRAREEEVNLDIIRAVFDGIEGSYVAENVE 301  
Db 289 KLLVLNPIKRGSLQIMKDRWMNVGH-----EEEEELK-----PYTEPPDP 329  
QY 302 RNDEGPLMMNAFEMITLS---QGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVA 358  
Db 330 FNDTKRIDI----MVTMGFARDEIN-DALINQKYDEVMATYILLGRKPP----- 373  
QY 359 NSMGFKSHTRNFKTRLEGLSSIKAGQL 385  
Db 374 -----EFEGGESLSSGNL 386

Search completed: October 20, 2003, 17:01:13  
Job time : 72 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:50:02 ; Search time 20 Seconds  
(without alignments)  
943.532 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTKKMRVGVGVGTIGEG.....IWRATEGIPKSEILRTITF 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	663	28.9	633	3	US-08-557-006C-43
2	662	28.9	552	3	US-08-557-006C-40
3	648	28.3	257	2	US-07-857-224B-25
4	643	28.0	345	3	US-09-101-146-1
5	636.5	27.8	257	3	US-09-101-146-6
6	591	25.8	793	4	US-09-523-849-32
7	584	25.5	729	2	US-08-677-298-2
8	584	25.5	729	4	US-09-523-849-33
9	579.5	25.3	779	4	US-08-817-832B-31
10	576.5	25.1	776	4	US-09-523-849-34
11	575	25.1	722	4	US-09-984-890-4
12	575	25.1	722	4	US-08-817-832B-32
13	575	25.1	745	4	US-09-523-849-36
14	574	25.0	668	4	US-09-930-181-2
15	574	25.0	724	4	US-09-984-890-2
16	561	24.5	604	4	US-09-523-849-35
17	543	23.7	339	3	US-08-688-988-33
18	523.5	22.8	351	3	US-08-688-988-28
19	511.5	22.3	631	4	US-09-579-664B-11
20	506	22.1	476	3	US-09-522-800-16
21	506	22.1	476	3	US-08-924-183-1
22	506	22.1	476	4	US-09-488-364-1
23	506	22.1	476	4	US-09-113-785-1
24	505.5	22.0	353	3	US-08-688-988-31
25	505.5	22.0	359	3	US-08-688-988-32
26	503.5	22.0	334	4	US-09-523-849-31
27	500.5	21.8	354	3	US-08-688-988-29

28	498.5	21.7	436	4	US-09-734-673-2	Sequence 2, Appli
29	498.5	21.7	436	4	US-09-523-849-2	Sequence 2, Appli
30	497	21.7	290	4	US-09-734-673-4	Sequence 4, Appli
31	492	21.5	476	3	US-08-924-183-2	Sequence 2, Appli
32	492	21.5	476	4	US-09-488-364-2	Sequence 2, Appli
33	490.5	21.4	363	3	US-08-688-988-30	Sequence 30, Appl
34	489.5	21.3	603	4	US-09-930-181-17	Sequence 17, Appl
35	487.5	21.3	252	2	US-07-857-224B-26	Sequence 26, Appl
36	487.5	21.3	496	3	US-08-924-183-10	Sequence 10, Appl
37	487.5	21.3	496	4	US-09-488-364-10	Sequence 10, Appl
38	487.5	21.3	496	4	US-09-113-785-3	Sequence 3, Appli
39	481	21.0	476	3	US-08-924-183-7	Sequence 7, Appli
40	481	21.0	476	4	US-09-488-364-7	Sequence 7, Appli
41	476.5	20.8	355	4	US-09-579-664B-10	Sequence 10, Appl
42	476	20.8	463	1	US-07-951-715A-25	Sequence 25, Appl
43	476	20.8	463	2	US-08-459-448A-25	Sequence 25, Appl
44	476	20.8	463	3	US-08-459-595A-25	Sequence 25, Appl
45	476	20.8	463	3	US-08-459-504B-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1  
US-08-557-006C-43  
; Sequence 43, Application US/08557006C  
; Patent No. 6258547  
; GENERAL INFORMATION:  
; APPLICANT: Beri, Rajindar K.  
; APPLICANT: Carling, David  
; APPLICANT: Forder, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING AMP-ACTIVATED PROTEIN KINASE  
; FILE REFERENCE: NGAP/PHM37588/UST  
; CURRENT APPLICATION NUMBER: US/08/557,006C  
; CURRENT FILING DATE: 1996-03-06  
; PRIOR APPLICATION NUMBER: PCT/GB94/01093  
; PRIOR FILING DATE: 1994-05-20  
; PRIOR APPLICATION NUMBER: GB 9310489.1  
; PRIOR FILING DATE: 1993-05-21  
; PRIOR APPLICATION NUMBER: GB 9318010.7  
; PRIOR FILING DATE: 1993-08-31  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 43  
; LENGTH: 633  
; TYPE: PRT  
; ORGANISM: Yeast  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)..(633)  
; OTHER INFORMATION: Yeast SNF1 polypeptide  
US-08-557-006C-43

Query Match 28.9%; Score 663; DB 3; Length 633;  
Best Local Similarity 39.4%; Pred. No. 1.7e-58;  
Matches 135; Conservative 80; Mismatches 98; Indels 30; Gaps 6;

QY	8	VGKYEVRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR	67
Db	52	IGNYQIVKTLGEGSGKVLAYHTTGQKVALKIINKKVLAKSDMQGRIEISYLRLLR	111
QY	68	HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIHVHKGRLSEESRKYFQQLVDVAHCHC	127
Db	112	HPIIKLYDVIKSKDEIIMVIEY-AGNELFDYIVQRDKMSEQEARFFQQIISAVEYCHR	170
QY	128	KGVHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGGQYD	187
Db	171	HKIVHRDLKPENLLDDEHLNVKIDFGLSNIMTDG-NFLKTCGSPNYAAPEVISGKLYA	229
QY	188	GSAADIWSCGVILFVILAGVLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDP	247
Db	230	GPEVDWSCGVILYVMLCRRLPFDDDESIPVLFKNISNGVVTLPKFLSPGAAGLIKMLIV	289

QY 248 NPKTRIQIQIKKDPWFRLN--YV---PIRAREEE-----VNLD 282  
Db 290 NPLNRISIHIMQDDWFKVDLPEYLLPPDLKPHPEEENENDSKKDSSPDNDEIDDLV 349

QY 283 DIRAVFDGIEGSYVAENVNRNDEGPL--MMNAFEMITLSQGL 322  
Db 350 NILSSTMGYEKDEIYESLESSEDTPAFNEIRDAYMLIKENKSL 392

RESULT 2  
US-08-557-006C-40  
; Sequence 40, Application US/08557006C  
; Patent No. 6258547  
; GENERAL INFORMATION:  
; APPLICANT: Beri, Rajindar K.  
; APPLICANT: Carling, David  
; APPLICANT: Forder, Robert A.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING AMP-ACTIVATED PROTEIN KINASE  
; FILE REFERENCE: NGAP/PHM37588/UST  
; CURRENT APPLICATION NUMBER: US/08/557,006C  
; CURRENT FILING DATE: 1996-03-06  
; PRIOR APPLICATION NUMBER: PCT/GB94/01093  
; PRIOR FILING DATE: 1994-05-20  
; PRIOR APPLICATION NUMBER: GB 9310489.1  
; PRIOR FILING DATE: 1993-05-21  
; PRIOR APPLICATION NUMBER: GB 9318010.7  
; PRIOR FILING DATE: 1993-08-31  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 40  
; LENGTH: 552  
; TYPE: PRT  
; ORGANISM: Rat  
; FEATURE:  
; NAME/KEY: gene  
; LOCATION: (1)..(1747)  
; OTHER INFORMATION: Full length cDNA sequence fragment of Human AMPK -  
; OTHER INFORMATION: fragment begins at nucleotide 24 and ends with  
; OTHER INFORMATION: nucleotide 1765  
US-08-557-006C-40

Query Match 28.9%; Score 662; DB 3; Length 552;  
Best Local Similarity 36.4%; Pred. No. 1.7e-58;  
Matches 149; Conservative 79; Mismatches 147; Indels 34; Gaps 7;  
QY 1 MTKKMR-----RVGKYEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQ 55  
Db 1 MAEQKHGDRVKIGHYVLGDTLGVGTGFKVIGEHQLTGHKVAVKILNRQKIRSLDVVGK 60  
QY 56 IKREISIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYP 115  
Db 61 IKREIQNLKLFRRPHIILKLYQVILSTPTDFFVMMEYVSGGELFDYICKHGRVEEVEARLF 120  
QY 116 QQLVDAVAHCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTENY 175  
Db 121 QQILSAVDYCHRHMVVHRDLKPENVLDDAQMNAKIADFGLSNMMSDG-EFLRTSCGSPNY 179  
QY 176 VAPEVLSQGQYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSA 235  
Db 180 AAPEVISGRLYAGPEVDIWSCGVILYALLCGTLPFDDDEHVPTLFKKIRGGVFIPEYLN 239  
QY 236 EVKFLIHRILDNPKNKTRIQIQIKKDPWFRNLNVYPIRAREE--EVNLD DIRAVFGIE- 292  
Db 240 SIATLLMHMLQVDPKLRATIKDIREHEWFKQDLPYLFDPDPSYDANVIDDEAVKEVCEK 299  
QY 293 ----GSYVAENVNRNDEGPLMMNAFEMITLSQGLNLSALFRRQDFVKRQTRFVSRREPS 348  
Db 300 FECTESEVMNSLYSGDPQDQLAVAYHLI-----IDNRRIMNQASEFYLAASPPT 348  
QY 349 EIIANIEAVANSMGFKSHTRNF-----KTR--LEGLSIKAGQLAV 387  
Db 349 GSEFDDMAMHIPPGLKPHPERMPPLIADSPKARCPLDALNTTKPSLAV 397

RESULT 3  
US-07-857-224B-25  
; Sequence 25, Application US/07857224B  
; Patent No. 5958784  
; GENERAL INFORMATION:  
; APPLICANT: Benner, Steven A.  
; TITLE OF INVENTION: Predicting Folded Structures of Proteins  
; NUMBER OF SEQUENCES: 114  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steven A. Benner  
; STREET: Hadlaubstrasse 151  
; CITY: Zurich  
; STATE: none  
; COUNTRY: Switzerland  
; ZIP: (note: this is an international post code) CH-8092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch diskette, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 7.0  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/857,224B  
; FILING DATE: 03/25/92  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA: none  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (International) 41 1 632 2830  
; TELEFAX: (International) 41 1 262 2437  
; TELEX: none  
; INFORMATION FOR SEQ ID NO: 25:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE:  
; DESCRIPTION: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Saccharomyces cerevisiae  
; FEATURE: Protein kinase; Table 8 Column 28  
; PUBLICATION INFORMATION:  
; AUTHORS:  
; AUTHORS: Hanks, S. K.  
; AUTHORS: Quinn, A. M.  
; AUTHORS: Hunter, T.  
; TITLE: The protein kinase family  
; JOURNAL: Science  
; VOLUME: 241  
; PAGES: 42-52  
; DATE: 1988  
US-07-857-224B-25

Query Match 28.3%; Score 648; DB 2; Length 257;  
Best Local Similarity 46.3%; Pred. No. 1.4e-57;  
Matches 120; Conservative 62; Mismatches 75; Indels 2; Gaps 2;  
QY 9 GKYEVGRTIGEGTFAKVKFARNTDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRH 68  
Db 1 GNYQIVKTLGEGSFGKVKLAYHTTTGQKVALKIINKKVLAKSDMQRIEREISYLRLLRH 60  
QY 69 PNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDAVAHCHCK 128  
Db 61 PHIILKYDVIKSDEIIMVIEY-AGNELFDYIVQRDKMSEQEARRFFQQLISAVEYCHRH 119  
QY 129 GVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSQGQYDG 188  
Db 120 KIVHRDLKPENLLDDEHLNVKIADFGLSNIMTDG-NFLKTS CGSPNYAAPEVISGKLYAG 178  
QY 189 SAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDPN 248  
Db 179 PEVDVWSCGVILYVMLCRRLLPFDDDESIPVLFKNISNGVYTLPKFLSPGAAGLIKRMILVN 238  
QY 249 PKTRIQIQIKKDPWFRLN 267

Db 239 PLNRISHEIMQDDWFKVD 257

RESULT 4

US-09-101-146-1  
; Sequence 1, Application US/09101146  
; Patent No. 6124125  
; GENERAL INFORMATION:  
; APPLICANT: Dartmouth College, St. Vincents Institute of  
; APPLICANT: Medical Research, Kemp et al.  
; TITLE OF INVENTION: No. 6124125el AMP Activated Protein Kinase  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jane Massey Licata, Esq.  
; STREET: 66 E. Main Street  
; CITY: Marlton  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: WINDOWS 95  
; SOFTWARE: WORDPERFECT 6.0 FOR WINDOWS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/101,146  
; FILING DATE: October 7, 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PN7450  
; FILING DATE: 8 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane Massey Licata  
; REGISTRATION NUMBER: 32,257  
; REFERENCE/DOCKET NUMBER: DC-0050  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (856) 810-1515  
; TELEFAX: (856) 810-1454  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345  
; TYPE: Amino acid  
; TOPOLOGY: Linear  
US-09-101-146-1

Query Match 28.0%; Score 643; DB 3; Length 345;  
Best Local Similarity 39.3%; Pred. No. 7e-57;  
Matches 136; Conservative 77; Mismatches 113; Indels 20; Gaps 6;  
Qy 1 MTKKVR---RVGKYEVGRITGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQI 56  
Db 1 NAEKQKHGRVKIGHYILGDTLGVGTGKVKVGKHELTGHKVAVKILNRQKI-RLDVGKI 59  
Qy 57 KREISIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKIFQ 116  
Db 60 RREIQNLKLFRRPHIILKLYQVISTPSDIFVMVEYVSGGELFDYICKNGRLDEKESRRLFQ 119  
Qy 117 QLVDAVAHCHCKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYV 176  
Db 120 QILSGVDYCHRMVVRDLKPENLVLLDAHNAKIADPGLSNMMSDG-EFLRTSCGSPNYA 178  
Qy 177 APEVLGGQYDGSAAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAEEFSCPPWFSAE 236  
Db 179 APEVLSGRLYAGPEVDIWSGVILYALLCGTLFPDDHVPVTLFKKICDGFYTPQYLNPS 238  
Qy 237 VKFLIHRILDENPKTRIQQIGIKKDPWFRNLVYPIRAREE-----EEVNLDDIRAVF 288  
Db 239 VISLLKHLQVDPKRAIKIDIREHEWFKQDLPKYLPEDPSPSYSTMIDDEALKEVCEKF 298  
Qy 289 DGIEGSYVAENVERNDEGFLMNAFEMITLSQGLNLSALFDRRQDF 334  
Db 299 ECSEEEVLSCLYRNHQDPLAV-AYHLI-----IDNRRIMNEAKDF 338

RESULT 5

US-09-101-146-6  
; Sequence 6, Application US/09101146  
; Patent No. 6124125  
; GENERAL INFORMATION:  
; APPLICANT: Dartmouth College, St. Vincents Institute of  
; APPLICANT: Medical Research, Kemp et al.  
; TITLE OF INVENTION: No. 6124125el AMP Activated Protein Kinase  
; NUMBER OF SEQUENCES: 64  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jane Massey Licata, Esq.  
; STREET: 66 E. Main Street  
; CITY: Marlton  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 08053  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE  
; COMPUTER: IBM PC  
; OPERATING SYSTEM: WINDOWS 95  
; SOFTWARE: WORDPERFECT 6.0 FOR WINDOWS  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/101,146  
; FILING DATE: October 7, 1998  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PN7450  
; FILING DATE: 8 JAN 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane Massey Licata  
; REGISTRATION NUMBER: 32,257  
; REFERENCE/DOCKET NUMBER: DC-0050  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (856) 810-1515  
; TELEFAX: (856) 810-1454  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257  
; TYPE: Amino acid  
; TOPOLOGY: Linear  
US-09-101-146-6

Query Match 27.8%; Score 636.5; DB 3; Length 257;  
Best Local Similarity 47.5%; Pred. No. 2e-56;  
Matches 122; Conservative 54; Mismatches 80; Indels 1; Gaps 1;  
Qy 9 GKYEVCRTIGEGTFAKVKFARNITDGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVRH 68  
Db 1 GHYILGDTLGVGTGKVKVGKHELTGHKVAVKILNRQKIRSLDVGKIRREIQNLKLFRRH 60  
Qy 69 PNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKIFQQLVDAVAHCHCK 128  
Db 61 PHILKLYQVISTPSDIFVMVEYVSGGELFDYICKNGRLDEKESRRRLFQQLSGVDYCHRH 120  
Qy 129 GYVHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYVAPVLSGQGYDG 188  
Db 121 MVVHRDLKPENLVLLDAHNAKIADPGLSNMMSDG-EFLRTSCGSPNYAAPEVISGRLYAG 179  
Qy 189 SAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAEEFSCPPWFSAEVKFLIHRILDPN 248  
Db 180 PEVDIWSGVILYALLCGTLFPDDHVPVTLFKKICDGFYTPQYLNPSVISLLKHLQVD 239  
Qy 249 PKTRIQQIGIKKDPWFR 265  
Db 240 PMKRATIKDIREHEWFK 256  
RESULT 6  
US-09-523-849-32  
; Sequence 32, Application US/09523849  
; Patent No. 6458561



```

; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scacheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, Dave
; TITLE OF INVENTION: HUMAN NIM1 KINASE
; FILE REFERENCE: PC-0009 US
; CURRENT APPLICATION NUMBER: US/09/523,849
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: GenBank Accession No. 6458561 g2052189
US-09-523-849-32
```

```

Query Match      25.8%; Score 591; DB 4; Length 793;
Best Local Similarity 35.1%; Pred. No. 4.8e-51;
Matches 135; Conservative 72; Mismatches 134; Indels 44; Gaps 7;

QY      2 TKMRRVGKYEVRTIGEGTFAKVKFARNDDTGDNVAIKIMAKSTILKNRMVDQIKREIS 61
Db      51 TDEQPHIGNYRLQKTIGKNFAKVKLARHVLGTREAVAVKIIDK-TQLNPTSLQKLFREVR 109

QY      62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVDA 121
Db      110 IMKILNHPNIVKLFVETIEKTLVLMVEYASGGEVFDYLVAHGRMKEARAKFRQIVSA 169

QY      122 VAHCHCKGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYVAPEVL 181
Db      170 VQYCHQKCIVHRDLKAENLLLDADNMNIADFGFSNEFTVG-NKLDFTCGSPPYAAPELF 228

QY      182 SGQGYDGSAAIWSGCVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLI 241
Db      229 QGKKYDGEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRVPFYMSTDCENLL 288

QY      242 HRILDNPKTRIQIQIKKDPWFRLNY--VPIRAREEEVENLDDIRAV----- 287
Db      289 KLLVLNPIKRGSLQIMKDRMNMNVGHEEELKPYSEPELDLNDAKRIDIMVTMGFARDE 348

QY      288 -----FDGIEGSYA-----ENVERNDEGPLMNAFEMITLSQGLNLSAL--- 327
Db      349 INDALVSQXYDEVMATYILLGRKPPPEGGESLSSG---NLCQSRPSSDLNNSTLQSP 404

QY      328 ----FDRRQDFVKRQTRFVSRREPS 348
Db      405 AHLKVQRSISANQKQRRFSDHAGPS 429
```

```

RESULT 7
US-08-677-298-2
; Sequence 2, Application US/08677298
; Patent No. 5863729
; GENERAL INFORMATION:
; APPLICANT: Piwnica-Worms, Helen
; TITLE OF INVENTION: DNA SEQUENCES ENCODING HUMAN TCAK-1
; TITLE OF INVENTION: KINASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80303
; COMPUTER READABLE FORM:
```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,298
; FILING DATE: 09-JUL-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Caruthers, Jennie M.
; REGISTRATION NUMBER: 34,464
; REFERENCE/DOCKET NUMBER: 9-96
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 729 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-677-298-2
```

```

Query Match      25.5%; Score 584; DB 2; Length 729;
Best Local Similarity 33.6%; Pred. No. 2.2e-50;
Matches 139; Conservative 81; Mismatches 154; Indels 40; Gaps 7;

QY      8 VGKYEVRTIGEGTFAKVKFARNDDTGDNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67
Db      53 IGNVRLTKTIGKGNFAKVKLARHILTGREVAIKIIDK-TQLNPTSLQKLFREVRIMKILN 111

QY      68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLESESRKYFQQLVDAVAHCHC 127
Db      112 HPNIVKLFVETIEKTLVLMVEYASGGEVFDYLVAHGRMKEARSKFRQIVSAVQYCHQ 171

QY      128 KGVYHRDLKPENLLDNTGNLKVSDPGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYD 187
Db      172 KRIVHRDLKAENLLLDADNMNIADFGFSNEFTVGK-LDTFCGSPPYAAPELFQKKYD 230

QY      188 GSAADIWSGCVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILD 247
Db      231 GPEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLLKRFLVL 290

QY      248 NPKTRIQIQIKKDPWFRLNYV--PIRAREEEVENLDDIRAV----- 287
Db      291 NPIKRGTLQIMKDRWINAGHDEDELKPFVEPELDISDQKRIDIMVGMGYSQEEIQESLS 350

QY      288 ---FDGIEGSYV-----AENVERNDEGPLMNAFEMITLSQGLNLSA-----LFDRRQD 333
Db      351 KMKYDEITATYLLGRKSSSELDASDSSSSNLSLAKVPSDDLNNSTGQSPHHKVRSVS 410

QY      334 FVKRQTRFVSRREPSEIITANIEAVANSMGFKSHTRNFKTRLEGLSSIKAGQLAV 387
Db      411 SSQKQRRYSDHAGP----AIPSWVAYPKRSQTSQTADGDLKEDGISSRKSSGSAV 460
```

```

RESULT 8
US-09-523-849-33
; Sequence 33, Application US/09523849
; Patent No. 6458561
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scacheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, Dave
; TITLE OF INVENTION: HUMAN NIM1 KINASE
; FILE REFERENCE: PC-0009 US
; CURRENT APPLICATION NUMBER: US/09/523,849
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 39
```

```
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GenBank Accession No. 6458561 g3089349
US-09-523-849-33

Query Match      25.5%; Score 584; DB 4; Length 729;
Best Local Similarity 33.6%; Pred. No. 2.2e-50;
Matches 139; Conservative 81; Mismatches 154; Indels 40; Gaps 7;

QY      8 VGKYEVRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67
Db      53 IGYRLKLTIGKGNFAKVKLARHILTGREVAIKIIDK-TQLNPTSLOKLFREVRIMKILN 111
QY      68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDVAHCHC 127
Db      112 HPNIVKLFVETETKTLYLIMEYASGGEVFDYLVAHGRMKEKEARSKFRQIVSAVOYCHQ 171
QY      128 KGVYHRDLKPENLLDTNGNLKVSDFGLSALPOEGVELLRTTCGTPNYVAPEVLSGQGYD 187
Db      172 KRIVHRDLKAENLLDADMNIKIADFGEFSNEFTVGGK-LDTFCGSPPPYAAPELFQGGKYD 230
QY      188 GSAADIWSCGVILFVILAGYLPFSETDPLGLYRKINAAEFSCPPWFSAEVKFLIHRILDP 247
Db      231 GPEVDVWSLGLVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLLKRFLVL 290
QY      248 NPKTRIQIQIKKDPWFRNLNV--PIRAREEEVNLDLDIRAV----- 287
Db      291 NPIKRGTLQIQIMKDRWINAGHEEDELKPFVEPELDISDQKRIDIMVGMGYSQEEIQESLS 350
QY      288 ---FDGIEGSYV-----AENVERNDEGPLMMNAFEMITLSQGLNLSA-----LFDRRQD 333
Db      351 KMKYDEITATYLLGLGRKSSELDASDSSSSNLSLAKVRPSSDLNNSGTGSPHHKVQRSVS 410
QY      334 FVKRQTRFVSRRPSEIIANIEAVANSMGFKSHTRNFKTRLEGLSIKAGQLAV 387
Db      411 SSQKQRRYS DHAGP----AIPSVVAYPKRSQTSTADGLKEDGISSRKS SGS AV 460
```

```
RESULT 9
US-08-817-832B-31
; Sequence 31, Application US/08817832B
; Patent No. 6579691
; GENERAL INFORMATION:
; APPLICANT: MANDELKOW, Eckhard, et al.
; TITLE OF INVENTION: No. 6579691el Protein Kinase (NPK-110)
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 S. Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: US
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,832B
; FILING DATE: 28-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP95/04258
; FILING DATE: 30-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94 11 7122.5
; FILING DATE: 28-OCT-1994
; INFORMATION FOR SEQ ID NO: 31:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 779 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-817-832B-31

Query Match      25.3%; Score 579.5; DB 4; Length 779;
Best Local Similarity 35.2%; Pred. No. 6.9e-50;
Matches 136; Conservative 70; Mismatches 135; Indels 45; Gaps 8;

QY      2 TKMRRVGVGYEVGRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREIS 61
Db      37 TDEQPHIGNYRLQXTIGKGNFAKVKLARHVLTGREVAVKIIDK-TQLNPTSLOKLFREVR 95
QY      62 IMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDA 121
Db      96 IMKILNHPNIVKLFVETETKTLVLMYASGGEVFDYLVAHGRMKEKEARAKFRQIVSA 155
QY      122 VAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPOEGVELLRTTCGTPNYVAPEVL 181
Db      156 VOYCHQKCTIVHRDLKAENLLDADMNIKIADFGEFSNEFTVG-NKLDTFCGSPPPYAAPELF 214
QY      182 SGQYDGSAAIWSGCVILFVILAGYLPFSETDPLGL-YRKINAAHFSCPPWFSAEVKFL 240
Db      215 QGKKYDGPVDVWSLGLVILYTLVSGSLPFDGQNLKELRERSCLRGKYRVFFYFYMSTDCENL 274
QY      241 IHRILDPNPKTRIQIQIKKDPWFRNLNV--VIPAREEEVNLDLDIRAV----- 287
Db      275 LKLLVLNLPKRGSLQIQIMKDRWMNVGHEEEELKPYSEPELDLNDAKRIDIMVTMGFARD 334
QY      288 -----FDGIEGSYVA-----ENVERNDEGPLMMNAFEMITLSQGLNLSAL-- 327
Db      335 EINDALVSQKYDEVMATYILLGRKPEFEFGESLSSG----NLCQSRSPSSDLNNSLTQS 390
QY      328 -----FDRQDFVKRQTRFVSRREPS 348
Db      391 PAHLKVQRTISANQKQRRFSDHAGPS 416
```

```
RESULT 10
US-09-523-849-34
; Sequence 34, Application US/095233849
; Patent No. 6458561
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Molteni, Angela
; APPLICANT: Magnaghi, Paola
; APPLICANT: Bosotti, Roberta
; APPLICANT: Scacheri, Emanuela
; APPLICANT: Isacchi, Antonella
; APPLICANT: Hodgson, Dave
; TITLE OF INVENTION: HUMAN NIM1 KINASE
; FILE REFERENCE: PC-0009 US
; CURRENT APPLICATION NUMBER: US/09/523,849
; CURRENT FILING DATE: 2000-03-13
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PERL Program
; SEQ ID NO 34
; LENGTH: 776
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GenBank Accession No. 6458561 g5672676
US-09-523-849-34
```

```
Query Match      25.1%; Score 576.5; DB 4; Length 776;
Best Local Similarity 34.3%; Pred. No. 1.4e-49;
Matches 146; Conservative 78; Mismatches 147; Indels 55; Gaps 11;

QY      3 KKMRRVGVGYEVGRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREISI 62
```

Db 19 QKPLRVGFYDVERTLGKGNFAVVKLARHRVTKTQVAIKIIDK-TRLDSSNLEKIYREVQL 77  
QY 63 MKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESESRKYFQQLVDVAV 122  
Db 78 MKLLNHPNIIKLYQVMTKMDMLYIVTEFAKNMGEMEDYLTSGHLSENEARKFWQILSAV 137  
QY 123 ARCHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLS 182  
Db 138 EYCHNHIVHRDLKTENLLLDGNMDIKLADFGFNGFYKPG-EPLSTWCGSPPYAAPVFE 196  
QY 183 GQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIH 242  
Db 197 GKEYEGPQLDIWSLGVVLYVLCGSLPFDGPNLPTLRQVLEGRFRIPFFMSQDCETLIR 256  
QY 243 RILDNPKTRIQIGIKKDPWFRNLNVVPIRAREEEV-----NLDD-----IRA 286  
Db 257 RMLVVDPAKRITIAQIRQHRWQAD--PTLLQQDDPAFSMQGYTSNLGDYNEQVLGIMQA 314  
QY 287 VFDGIEGSVVAENVERDEGLPMNAFEMITLSQGLNLSALFDRRQDFVKFQTRFVSRRE 346  
Db 315 L--GIDRQRTVESLQNSS-----YNHFAAI-----YLLLERLRE--HRSTQPSSRAT 358  
QY 347 PSEIIANIEAVANSMGFKSHTRNFKTRLEGSLSKAGQLAVVIEIYEVAPSLFMVDVRKA 406  
Db 359 PAP-----ARQPQLRNSDLSSLEVPQEILPCDPFR--PSLLCPQPQAL 399  
QY 407 AGETLE 412  
Db 400 AQSVLQ 405

RESULT 11  
US-09-984-890-4  
; Sequence 4, Application US/09984890  
; Patent No. 6492156  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001306  
; CURRENT APPLICATION NUMBER: US/09/984,890  
; CURRENT FILING DATE: 2001-10-31  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 722  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
US-09-984-890-4

Query Match 25.1%; Score 575; DB 4; Length 722;  
Best Local Similarity 41.9%; Pred. No. 1.8e-49;  
Matches 114; Conservative 61; Mismatches 89; Indels 8; Gaps 3;  
QY 8 VGKYEVRTTIGEGTFAKVKFARNTTDGDNVVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67  
Db 50 IGYRLTKTIGKGNFAKVKLARHILTGKEVAVKIIDK-TQLNSSSLOKLFREVIRMKVLN 108  
QY 68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESESRKYFQQLVDAVAHCHC 127  
Db 109 HPNIVKLFEVIEKTLTYLVMEYASGGEVFDYLVAHGRMKEKEARAKFRQIVSAVOYCHQ 168  
QY 128 KGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYD 187  
Db 169 KFIVHRDLKAENLLDADMNIKIADFGFSNEFTFG-NKLDTFCGSPPYAAPELFQKKYD 227  
QY 188 GSAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDP 247  
Db 228 GPEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLLKKFLIL 287  
QY 248 NPKTRIQIGIKKDPWFRNLNVVPIRAREEEV 279

Db 288 NPSKRGTLQIMKDRWMNVGH-----EDDEL 313  
RESULT 12  
US-08-817-832B-32  
; Sequence 32, Application US/08817832B  
; Patent No. 6579691  
; GENERAL INFORMATION:  
; APPLICANT: MANDELKOW, Eckhard, et al.  
; TITLE OF INVENTION: No. 6579691el Protein Kinase (NPK-110)  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 S. Wacker Drive, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: US  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,832B  
; FILING DATE: 28-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: WO PCT/EP95/04258  
; FILING DATE: 30-OCT-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 94 11 7122.5  
; FILING DATE: 28-OCT-1994  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 722 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-817-832B-32

Query Match 25.1%; Score 575; DB 4; Length 722;  
Best Local Similarity 41.9%; Pred. No. 1.8e-49;  
Matches 114; Conservative 61; Mismatches 89; Indels 8; Gaps 3;  
QY 8 VGKYEVRTTIGEGTFAKVKFARNTTDGDNVVAIKIMAKSTILKNRMVDQIKREISIMKIVR 67  
Db 50 IGYRLTKTIGKGNFAKVKLARHILTGKEVAVKIIDK-TQLNSSSLOKLFREVIRMKVLN 108  
QY 68 HPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESESRKYFQQLVDAVAHCHC 127  
Db 109 HPNIVKLFEVIEKTLTYLVMEYASGGEVFDYLVAHGRMKEKEARAKFRQIVSAVOYCHH 168  
QY 128 KGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQGYD 187  
Db 169 KFIVHRDLKAENLLDADMNIKIADFGFSNEFTFG-NKLDTFCGSPPYAAPELFQKKYD 227  
QY 188 GSAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRILDP 247  
Db 228 GPEVDVWSLGVILYTLVSGSLPFDGQNLKELRERVLRGKYRIPFYMSTDCENLLKKFLIL 287  
QY 248 NPKTRIQIGIKKDPWFRNLNVVPIRAREEEV 279  
Db 288 NPSKRGTLQIMKDRWMNVGH-----EDDEL 313

RESULT 13  
US-09-523-849-36  
; Sequence 36, Application US/09523849  
; Patent No. 6458561  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 20, 2003, 16:45:11 ; Search time 44 Seconds  
(without alignments)  
1608.910 Million cell updates/sec

Title: US-09-824-735-2  
Perfect score: 2293  
Sequence: 1 MTKMRRVKGVEVGTIGEG.....IWRATEGIPKSEILRTTF 446

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length DB	ID	Description
1	2293	100.0	446	21 AAG50962	Arabidopsis thalia
2	2293	100.0	446	23 ABG31356	Arabidopsis thalia
3	2288	99.8	446	23 ABG31357	Arabidopsis thalia
4	2285	99.7	446	23 ABG31358	Arabidopsis thalia
5	2273	99.1	442	21 AAG50963	Arabidopsis thalia
6	2085	90.9	405	21 AAG50964	Arabidopsis thalia
7	1484.5	64.7	445	21 ABR40808	Arabidopsis thalia
8	1484.5	64.7	445	24 ABR40808	Arabidopsis thalia
9	1469	64.1	481	21 AAB03415	Corn putative carb

10	1469	64.1	481	24	ABR40707	Zea mays oil trait
11	1367	59.6	441	21	AAB03423	Soybean putative c
12	1367	59.6	441	24	ABR40717	Glycine max oil tr
13	1344.5	58.6	416	21	AAG14342	Arabidopsis thalia
14	1336.5	58.3	409	21	AAG31349	Arabidopsis thalia
15	1314	57.3	441	24	ABR40817	Arabidopsis thalia
16	1303.5	56.8	396	21	AAG31350	Arabidopsis thalia
17	1288.5	56.2	441	22	AAB85211	P. patens protein
18	1196.5	52.2	380	21	AAG14343	Arabidopsis thalia
19	1163.5	50.7	367	21	AAG14344	Arabidopsis thalia
20	1063	46.4	422	21	AAB03419	Soybean putative c
21	1063	46.4	422	24	ABR40713	Glycine max oil tr
22	1043	45.5	438	21	AAB03421	Soybean putative c
23	1043	45.5	438	24	ABR40715	Glycine max oil tr
24	1040	45.4	441	24	ABR40814	Arabidopsis thalia
25	1039.5	45.3	439	21	AAG39523	Arabidopsis thalia
26	1039.5	45.3	439	21	AAG39523	Arabidopsis thalia
27	997.5	43.5	332	21	AAG52661	Arabidopsis thalia
28	970.5	42.3	312	21	AAG52662	Arabidopsis thalia
29	970.5	42.3	374	21	AAG34603	Arabidopsis thalia
30	932	40.6	380	21	AAG26367	Arabidopsis thalia
31	930	40.6	366	21	AAG26368	Arabidopsis thalia
32	911	39.7	416	21	AAG06699	Arabidopsis thalia
33	911	39.7	433	21	AAG06698	Arabidopsis thalia
34	876.5	38.2	376	21	AAG39525	Arabidopsis thalia
35	849.5	37.0	273	21	AAG52663	Arabidopsis thalia
36	849.5	37.0	335	21	AAG34604	Arabidopsis thalia
37	835	36.4	351	21	AAG49615	Arabidopsis thalia
38	833	36.3	337	21	AAG49616	Arabidopsis thalia
39	809.5	35.3	324	21	AAG34605	Arabidopsis thalia
40	808.5	35.3	371	21	AAG54418	Zea mays protein f
41	808.5	35.3	431	21	AAG48083	Arabidopsis thalia
42	808.5	35.3	455	21	AAG48082	Arabidopsis thalia
43	775	33.8	307	21	AAG26369	Arabidopsis thalia
44	757	33.0	353	21	AAG06700	Arabidopsis thalia
45	719	31.4	523	21	AAB03425	Wheat putative car

ALIGNMENTS

RESULT 1  
AAG50962  
ID AAG50962 standard; Protein; 446 AA.  
XX  
AC AAG50962;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64634.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX

PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.



PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 100.0%; Score 2293; DB 21; Length 446;  
Best Local Similarity 100.0%; Pred. No. 5e-237;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKMRRVGKYEVRTIGGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDQIKREI 60  
Db 1 MTKMRRVGKYEVRTIGGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDQIKREI 60  
QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVD 120  
Db 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVD 120  
QY 121 AVAHCHCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEV 180  
Db 121 AVAHCHCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEV 180  
QY 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAEEFSCPPWFSAEVKFL 240  
Db 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAEEFSCPPWFSAEVKFL 240  
QY 241 IHRILDPNPKTRIQQIGIKKDPWFRNLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
Db 241 IHRILDPNPKTRIQQIGIKKDPWFRNLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
QY 301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
Db 301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
QY 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEYVAPSLFMVDVRKAAGETLEYHKFYKKL 420  
Db 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEYVAPSLFMVDVRKAAGETLEYHKFYKKL 420  
QY 421 CSKLENIINWATEGIPKSEILRTITF 446  
Db 421 CSKLENIINWATEGIPKSEILRTITF 446

RESULT 2  
ABG31356  
ID ABG31356 standard; Protein; 446 AA.  
XX ABG31356;  
XX AC  
XX XX  
DT 15-NOV-2002 (first entry)  
XX Arabidopsis thaliana SOS2 serine/threonine protein kinase.  
DE Arabidopsis thaliana SOS2 serine/threonine protein kinase.  
XX Salt overly sensitive 7; SOS2; serine/threonine protein kinase;  
KW

KW salt tolerance; agricultural crop; rice; corn; wheat; cotton;  
KW peanut; soybean; plant protectant; plant; transgenic.  
XX Arabidopsis thaliana.  
OS Arabidopsis thaliana.  
XX US2002095032-A1.  
XX 18-JUL-2002.  
XX 04-APR-2001; 2001US-0824735.  
XX 04-APR-2000; 2000US-194649P.  
XX (UYAR-) UNIV ARIZONA.  
PI Zhu J, Liu J, Ishitani M, Kim C, Halfter U;  
XX WPI; 2002-665801/71.  
DR N-PSDB; ABK91073, ABK91084.  
XX New SOS2 protein kinase polypeptide and nucleic acids encoding the  
PT polypeptide, useful for increasing salt tolerance to plants, especially  
PT to agricultural crops -  
XX Claim 37; Fig 2; 28pp; English.

PS The present invention relates to the isolation of an Arabidopsis  
XX thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine  
CC protein kinase. The polynucleotide sequence encoding SOS2 can be used  
CC to increase salt tolerance in plants, particularly agricultural crops  
CC such as rice, corn, wheat, cotton, peanut and soybean. The  
CC polynucleotide sequence encoding SOS2 can also be used to generate  
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to  
CC isolate those cDNAs or genes which exhibit a high degree of similarity  
CC to the sequence of the SOS2 gene, and as PCR primers for the production  
CC of DNA which encodes an enzyme having serine/threonine protein kinase  
CC activity. The present sequence represents Arabidopsis thaliana SOS2  
CC serine/threonine protein kinase.  
XX

SQ Sequence 446 AA;

Query Match 100.0%; Score 2293; DB 23; Length 446;  
Best Local Similarity 100.0%; Pred. No. 5e-237;  
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTKMRRVGKYEVRTIGGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDQIKREI 60  
Db 1 MTKMRRVGKYEVRTIGGTFAKVKFARNTDGNVAIKIMAKSTILKNRMVDQIKREI 60  
QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVD 120  
Db 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYYFQQLVD 120  
QY 121 AVAHCHCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEV 180  
Db 121 AVAHCHCKGVYHRDLKPNLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEV 180  
QY 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAEEFSCPPWFSAEVKFL 240  
Db 181 LSGQGYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAEEFSCPPWFSAEVKFL 240  
QY 241 IHRILDPNPKTRIQQIGIKKDPWFRNLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
Db 241 IHRILDPNPKTRIQQIGIKKDPWFRNLNYPVIRAREEEVNLDLDIRAVFDGIEGSYVAENV 300  
QY 301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
Db 301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360  
QY 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEYVAPSLFMVDVRKAAGETLEYHKFYKKL 420  
Db 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEYVAPSLFMVDVRKAAGETLEYHKFYKKL 420



```
CC polynucleotide sequence encoding SOS2 can also be used to generate
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to
CC isolate those cDNAs or genes which exhibit a high degree of similarity
CC to the sequence of the SOS2 gene, and as PCR primers for the production
CC of DNA which encodes an enzyme having serine/threonine protein kinase
CC activity. The present sequence represents a mutant of Arabidopsis
CC thaliana SOS2 serine/threonine protein kinase.
CC Note: The present sequence is not given in the specification but
CC is created by the indexer from the information given on page 5.
XX
SQ Sequence 446 AA;

Query Match 99.7%; Score 2285; DB 23; Length 446;
Best Local Similarity 99.8%; Pred. No. 3.6e-236;
Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKMRRVGKYEVGRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREI 60
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
1 MTKMRRVGKYEVGRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREI 60
QY 61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVD 120
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
61 SIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVD 120
QY 121 AVAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPOEGVELLRITTCGTPNYVAPEV 180
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
121 AVAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPOEGVELLRITTCGTPNYVAPEV 180
QY 181 LSGQYDGSAAIWSGCVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
181 LSGQYDGSAAIWSCEVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFL 240
QY 241 IHRILDPNPKTRIQQIKKDPWFRNLNYPTRAREEEVNLLDIRAVFDGIEGSYVAENV 300
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
241 IHRILDPNPKTRIQQIKKDPWFRNLNYPTRAREEEVNLLDIRAVFDGIEGSYVAENV 300
QY 301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
301 ERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANS 360
QY 361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMDVVRKAAGTLEYHKFYKKL 420
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
361 MGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMDVVRKAAGTLEYHKFYKKL 420
QY 421 CSKLENIWRATEGIPKSEILRTITF 446
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
421 CSKLENIWRATEGIPKSEILRTITF 446

RESULT 5
AAG50963
ID AAG50963 standard; Protein; 442 AA.
XX
AC AAG50963;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 64635.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR
```

```
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
```



PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 99.1%; Score 2273; DB 21; Length 442;  
Best Local Similarity 100.0%; Pred. No. 7e-235;  
Matches 442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 MRRVGKYEVRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREISIMK 64  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
1 MRRVGKYEVRTIGEGTFAKVKFARNTTDGNVAIKIMAKSTILKNRMVDQIKREISIMK 60  
  
QY 65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESRKYFQQLVDAVAH 124  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
61 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESRKYFQQLVDAVAH 120  
  
QY 125 CHCKGVYHRDLKPENLLDDTNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQ 184  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
121 CHCKGVYHRDLKPENLLDDTNGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQ 180  
  
QY 185 GYDGSADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
181 GYDGSADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 240  
  
QY 245 LDPNPKTRIQQIGIKKDPWFRNLNVPIRAREEEEEVNLLDDIRAVFDGIEGSYVAENVERND 304  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
241 LDPNPKTRIQQIGIKKDPWFRNLNVPIRAREEEEEVNLLDDIRAVFDGIEGSYVAENVERND 300  
  
QY 305 EGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGFK 364  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
301 EGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVANSMGFK 360  
  
QY 365 SHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDRVKAAGETLEYHKFKLCSKL 424  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
361 SHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMVDRVKAAGETLEYHKFKLCSKL 420  
  
QY 425 ENIIWRATEGIPKSEILRTITF 446  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
421 ENIIWRATEGIPKSEILRTITF 442

RESULT 6  
AAG50964  
ID AAG50964 standard; Protein; 405 AA.

[illegible]

PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 90.9%; Score 2085; DB 21; Length 405;  
Best Local Similarity 100.0%; Pred. No. 9.8e-215;  
Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 42 MAKSTILKNRMVDQIKREISIMKIIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIV 101  
Db 1 MAKSTILKNRMVDQIKREISIMKIIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIV 60  
QY 102 HKGRLEESESRYFQQLVDAVAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQE 161  
Db 61 HKGRLEESESRYFQQLVDAVAHCHCKGVYHRDLKPENLLDTNGNLKVSDFGLSALPQE 120  
QY 162 GVELLRITTCGTPNVVAPVLSGGQYDGSADIWSCGVILFVILAGVLPFSETDLPGLYRK 221  
Db 121 GVELLRITTCGTPNVVAPVLSGGQYDGSADIWSCGVILFVILAGVLPFSETDLPGLYRK 180  
QY 222 INAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQQIKKDPWFRLNVVPIRAREEEVN 281  
Db 181 INAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQQIKKDPWFRLNVVPIRAREEEVN 240

QY 282 DDIRAVFDGIEGSYVAENVERNDEGPLMNAFEMITLSQGLNLSALFDRQDFVKRQTRF 341  
Db 241 DDIRAVFDGIEGSYVAENVERNDEGPLMNAFEMITLSQGLNLSALFDRQDFVKRQTRF 300  
QY 342 VSRREPSEIIANIEAVANSMGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMV 401  
Db 301 VSRREPSEIIANIEAVANSMGFKSHTRNFKTRLEGLSSIKAGQLAVVIEIYEVAPSLFMV 360  
QY 402 DVRKAAGETLEYHKFYKKLCSKLENIWIRATEGIPKSEILRTITF 446  
Db 361 DVRKAAGETLEYHKFYKKLCSKLENIWIRATEGIPKSEILRTITF 405

RESULT 7  
AAG31348  
ID AAG31348 standard; Protein; 445 AA.  
XX  
AC AAG31348;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 376332.

KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.

OS Arabidopsis thaliana.

XX  
PN EP1033405-A2.

XX  
PD 06-SEP-2000.

XX  
PF 25-FEB-2000; 2000EP-0301439.

XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.



PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 64.7%; Score 1484.5; DB 21; Length 445;  
Best Local Similarity 62.3%; Pred. No. 4e-150;

Matches 273; Conservative 84; Mismatches 80; Indels 1; Gaps 1;	
QY	5 MRRVGKYEVRTIGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMK 64
Db	3 VRKVGKYEELGRTIGEGTFAKVKFAQNTETGESVAMKIVDRSTIIKRKMVDQIKREISIMK 62
QY	65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDAVAH 124
Db	63 LVRHPCVRLYEVLASRTKIYIILEYITGGELFDKIVRNGRLSEBARKYFHQLIDGVY 122
QY	125 CHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQ 184
Db	123 CHSKGVYHRDLKPENLLDSSQGNLKISDFGLSALPEQGVTLKTTCTGTPNYVAPEVLSHK 182
QY	185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244
Db	183 GYNGAVADIWSCGVILYVLMAGYLPFDEMDLPTLYSKIDKAEFSCPSYFALGAKSLINRI 242
QY	245 LDPNPKTRIQQIKDPWFRNLNYVPIRAREEEEEVNLDDIRAVFDGIEGSYVAENVERND 304
Db	243 LDPNPETRIITAEIRKDEWFLKDYTPVQLIDYEHVNLDDVYAAFDDEEQTYAQDGT-R 301
QY	305 EGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTFVSRREPSEIIANIEAVANSMGFK 364
Db	302 TGPLTLNADFLLIILSQGLNLATLFDGRKDSMKHQTRFISHKPANVVLSSMEVVSQSMGFK 361
QY	365 SHTRNFKTRLEGLSSIKAGQLAWVIEIYEVAPSLFMVDVRKAAGETLEYHKFKYKLC SKL 424
Db	362 THIRNYKMRVEGLSANKTSHFSVILEVFKVAPSILMVDIQNAAGDAEAYLKFKYKTFCSKL 421
QY	425 ENIIWRATEGIPKSEILR 442
Db	422 DDIIWKPPDASMRNRVTK 439
RESULT 8	
ABR40808	
ID	ABR40808 standard; Protein; 445 AA.
XX	AC ABR40808;
XX	16-MAY-2003 (first entry)
DE	Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:400.
KW	Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;
KW	receptor-like protein kinase; mitogen activated protein kinase; oil;
KW	LIP15-like transcription factor caleosin; ATP citrate lyase; SNF1;
KW	CKC-like transcription factor; antisense inhibition; co-suppression;
XX	transgenic plant.
OS	Arabidopsis thaliana.
XX	WO2003002751-A2.
PF	09-JAN-2003.
XX	27-JUN-2002; 2002WO-US20152.
XX	29-JUN-2001; 2001US-301913P.
PA	(DUPO ) DU PONT DE NEMOURS & CO E I.
PA	(PION-) PIONEER HI-BRED INT INC.
XX	Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;
PI	Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H;
PI	Shen B, Tarczynski MC;
XX	WPI; 2003-201509/19.
DR	Novel nucleotide fragment encoding polypeptides having receptor-like
XX	protein kinase activity, caleosin-like activity, useful for altering
PT	oil phenotypes in plants such as sunflower, coconut, soybean, wheat and
PT	
PT rice -	
XX	Claim 12; Page 431-432; 542pp; English.
PS	The present invention describes an isolated nucleotide fragment (I)
XX	comprising a nucleic acid sequence (NS) chosen from a NS encoding a
CC	polypeptide (pp) having receptor-like protein kinase activity, mitogen
CC	activated protein (MAP)-kinase activity, LIP15-like transcription factor
CC	activity, caleosin-like activity, ATP citrate lyase activity, SNF1-like
CC	activity and CKC-like transcription factor activity. Also described:
CC	(1) complement (II) of (I); (2) a chimeric construct (III) comprising (I)
CC	or (II), operably linked to a regulatory sequence; (3) a plant (IV)
CC	comprising (III) in its genome; (4) seeds (V) obtained from (IV); and
CC	(5) oil obtained from (V). (I) or its part can be used in antisense
CC	inhibition or co-suppression in a transformed plant. (III) is useful for
CC	altering the oil phenotype in a plant such as corn, soybean, wheat, rice,
CC	canola, Brassica, sorghum, sunflower or coconut. (III) is also useful for
CC	creating transgenic plants having altered lipid profiles. (I) can also
CC	be used as a hybridisation probe. ACC00626 to ACC00868 and ABR40591 to
CC	ABR40879 represent sequences used in the exemplification of the present
CC	invention.
XX	
SQ	Sequence 445 AA;
Query Match	
Best Local Similarity 64.7%; Score 1484.5; DB 24; Length 445;	
Matches 273; Conservative 84; Mismatches 80; Indels 1; Gaps 1;	
QY	5 MRRVGKYEVRTIGEGTFAKVKFARNTDGDNVAIKIMAKSTILKNRMVDQIKREISIMK 64
Db	3 VRKVGKYEELGRTIGEGTFAKVKFAQNTETGESVAMKIVDRSTIIKRKMVDQIKREISIMK 62
QY	65 IVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFDRIVHKGRLEESERKYFQQLVDAVAH 124
Db	63 LVRHPCVRLYEVLASRTKIYIILEYITGGELFDKIVRNGRLSEBARKYFHQLIDGVY 122
QY	125 CHCKGVYHRDLKPENLLDNTGNLKVSDFGLSALPQEGVELLRTTCGTPNYVAPEVLSGQ 184
Db	123 CHSKGVYHRDLKPENLLDSSQGNLKISDFGLSALPEQGVTLKTTCTGTPNYVAPEVLSHK 182
QY	185 GYDGSAAADIWSCGVILFVILAGYLPFSETDLPGLYRKINAAEFSCPPWFSAEVKFLIHRI 244
Db	183 GYNGAVADIWSCGVILYVLMAGYLPFDEMDLPTLYSKIDKAEFSCPSYFALGAKSLINRI 242
QY	245 LDPNPKTRIQQIKDPWFRNLNYVPIRAREEEEEVNLDDIRAVFDGIEGSYVAENVERND 304
Db	243 LDPNPETRIITAEIRKDEWFLKDYTPVQLIDYEHVNLDDVYAAFDDEEQTYAQDGT-R 301
QY	305 EGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQTFVSRREPSEIIANIEAVANSMGFK 364
Db	302 TGPLTLNADFLLIILSQGLNLATLFDGRKDSMKHQTRFISHKPANVVLSSMEVVSQSMGFK 361
QY	365 SHTRNFKTRLEGLSSIKAGQLAWVIEIYEVAPSLFMVDVRKAAGETLEYHKFKYKLC SKL 424
Db	362 THIRNYKMRVEGLSANKTSHFSVILEVFKVAPSILMVDIQNAAGDAEAYLKFKYKTFCSKL 421
QY	425 ENIIWRATEGIPKSEILR 442
Db	422 DDIIWKPPDASMRNRVTK 439
RESULT 9	
AAB03415	
ID	AAB03415 standard; Protein; 481 AA.
XX	AC AAB03415;
XX	03-JAN-2001 (first entry)
DE	Corn putative carbon catabolite repression protein SNF1 #1.
XX	Corn; carbon catabolite repression; sucrose non-fermenting protein 1;
KW	SNF1; plant growth.
KW	











QY	245	LDPNPKTRIQIGIKKDPWFRLNYPVIRAREEEVNLDIDRAVFDGIEGYSVAENV	300
Db	243	LDPNPETRITIAEIRKDEWFLKDYTPQLIDYEHVNLDVYAADFDDPEQTYAQDGR-D	301
QY	305	EGPLMNAFEMITLSQGLNLSALFDRRQDFVKRQTRFVSRREPSEIIANIEAVNSMGFK	364
Db	302	TGPLTLNAFDLIILSQGLNLATLFDRGKDSMKHQTRFISHKPANVVLLSMEVVQS	361
QY	365	SHTRNFKTRLEGLSSIKAGQLAVVIE	390
Db	362	THIRNYKMRVEGLSANKTSHFSVILE	387
RESULT 14			
AAG31349			
ID	AAG31349 standard; Protein; 409 AA.		
XX	AAG31349;		
XX			
DT	17-OCT-2000 (first entry)		
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 37633.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway;		
KW	hybridisation assay; genetic mapping; gene expression control; promoter;		
KW	termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999; 99US-0121825.		
PR	05-MAR-1999; 99US-0123180.		
PR	09-MAR-1999; 99US-0123548.		
PR	23-MAR-1999; 99US-0125788.		
PR	25-MAR-1999; 99US-0126264.		
PR	29-MAR-1999; 99US-0126785.		
PR	01-APR-1999; 99US-0127462.		
PR	06-APR-1999; 99US-0128234.		
PR	08-APR-1999; 99US-0128714.		
PR	16-APR-1999; 99US-0129845.		
PR	19-APR-1999; 99US-0130077.		
PR	21-APR-1999; 99US-0130449.		
PR	23-APR-1999; 99US-0130510.		
PR	23-APR-1999; 99US-0130891.		
PR	28-APR-1999; 99US-0131449.		
PR	30-APR-1999; 99US-0132048.		
PR	30-APR-1999; 99US-0132407.		
PR	04-MAY-1999; 99US-0132484.		
PR	05-MAY-1999; 99US-0132485.		
PR	06-MAY-1999; 99US-0132486.		
PR	06-MAY-1999; 99US-0132487.		
PR	07-MAY-1999; 99US-0132863.		
PR	11-MAY-1999; 99US-0134256.		
PR	14-MAY-1999; 99US-0134218.		
PR	14-MAY-1999; 99US-0134219.		
PR	14-MAY-1999; 99US-0134221.		
PR	14-MAY-1999; 99US-0134370.		
PR	18-MAY-1999; 99US-0134768.		
PR	19-MAY-1999; 99US-0134941.		
PR	20-MAY-1999; 99US-0135124.		
PR	21-MAY-1999; 99US-0135353.		
PR	24-MAY-1999; 99US-0135629.		
PR	25-MAY-1999; 99US-0136021.		
PR	27-MAY-1999; 99US-0136392.		
PR	28-MAY-1999; 99US-0136782.		
PR	01-JUN-1999; 99US-0137222.		
PR	03-JUN-1999; 99US-0137528.		



PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 58.3%; Score 1336.5; DB 21; Length 409;  
Best Local Similarity 60.9%; Pred. No. 2.8e-134;

Matches 246; Conservative 77; Mismatches 80; Indels 1; Gaps 1;  
QY 39 IKIMAKSTILKNRMVDQIKREISIMKIVRHPNIVRLYEVLASPSKIYIVLEFVTGGELFD 98  
Db :||: :|||:| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:||||| 60  
1 MKIVDRSTIIKRKMVDQIKREISIMKLVRRHPCVVRLYEVLASRTKIYIILEYITGGELFD  
QY 99 RIVHKGRLESESRKYFQOLVDVAHCHCKGVYHRDLKPENLLDFTNGNLKVSDFGLSAL 158  
Db :||| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| 120  
61 KIVRNGRLSESEARKYFHQLIDGVDYCHSKGVYHRDLKPENLLDLSQGNLKISDFGLSAL  
QY 159 PQEGVELLRTTCGTPNYVAPEVLSGQYDGSAAADIWSCGVILFVILAGYLPFSETDLPGL 218  
Db ||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| 180  
121 PEQGVTLTKTCGTPNYVAPEVLSHKYNGAVADIWSCGVILYVLMAGYLPFDEMDLPTL  
QY 219 YRKINAAEFSCPPWFSAEVKFLIHRILDPNPKTRIQIQIKKDPWFRNLNVVPIRAREEEE 278  
Db ||:| :|||:| :||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| 240  
181 YSKIDKAEFSCPSYFALGAKSLINRILDPNPETRITIAEIRKDEWFLKDYTPVQLIDYEH  
QY 279 VNLLDIRAVFDGIEGSYVAENVVERNDEGPLMMNAFEMITLSQGLNLSALFDRRQDFVKRQ 338  
Db ||||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| 299  
241 VNLLDVYAAPDDPEEQTYAQDGT-R-DTGPLTLNAFDLIILSQGLNLATLFDRGKDSMKHQ  
QY 339 TRFVSRREPSEILIANIEAVANSMGFKSHTRNFKTRLEGSLSSIKAGQLAVVIEIYEVAPSL 398  
Db ||||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| 359  
300 TRFISHKPANVVVLSMEVVSQSMGFKTHIRNYKMRVEGLSANKTSHFSVILEVFKVAPSI  
QY 399 FMVDVRKAAGETLEYHKFKKLCSKLENIWRATEGIPKSEILR 442  
Db ||||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:| :|||:|  
360 LMVDIQNAAGDAEAYLKFKYKTFCSKLDLDIWKPPDASMRNRVTK 403  
RESULT 15  
ABR40817  
ID ABR40817 standard; Protein; 441 AA.  
XX  
AC ABR40817;  
XX  
DT 16-MAY-2003 (first entry)  
XX  
DE Arabidopsis thaliana oil trait related protein sequence SEQ ID NO:409.  
XX  
KW Plant; oil trait; oil phenotype; altered lipid profile; MAP kinase;  
KW receptor-like protein kinase; mitogen activated protein kinase; oil;  
KW LIP15-like transcription factor caleosin; ATP citrate lyase; SNF1;  
KW CKC-like transcription factor; antisense inhibition; co-suppression;  
transgenic plant.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO2003002751-A2.  
XX  
PD 09-JAN-2003.  
XX  
PF 27-JUN-2002; 2002WO-US20152.  
XX  
PR 29-JUN-2001; 2001US-301913P.  
XX  
PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX (PION-) PIONEER HI-BRED INT INC.  
PI Allen SM, Allen WB, Cahoon RE, Epelbaum S, Famodu OO, Harvell LT;  
PI Jones TJ, Kinney AJ, Klein TM, Li C, Oliveira IC, Sakai H;  
PI Shen B, Tarczynski MC;  
XX  
DR WPI; 2003-201509/19.  
XX  
PT Novel nucleotide fragment encoding polypeptides having receptor-like  
protein kinase activity, caleosin-like activity, useful for altering  
oil phenotypes in plants such as sunflower, coconut, soybean, wheat and  
rice -  
XX  
PS Claim 12; Page 446-447; 542pp; English.  
XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 01:18:57 ; Search time 18159 Seconds  
(without alignments)  
11588.695 Million cell updates/sec

Title: US-09-824-735-1  
Perfect score: 5144  
Sequence: 1 tttttggcggaataatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*

- 1: gb\_ba:\*
- 2: gb\_htg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_sts:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htg\_hum:\*
- 31: em\_htg\_inv:\*
- 32: em\_htg\_other:\*
- 33: em\_htg\_mus:\*
- 34: em\_htg\_pln:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_mam:\*
- 37: em\_htg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_htgo\_hum:\*
- 40: em\_htgo\_mus:\*
- 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5144	100.0	5144	8	AF237670	AF237670 Arabidops
2	5144	100.0	50015	8	AB025611	AB025611 Arabidops
3	4527	88.0	16228	8	F6I13	AF262044 Arabidops
C 4	459.4	8.9	110469	8	AC006258	AC006258 Arabidops
C 5	459.4	8.9	110680	8	AC006259	AC006259 Arabidops
C 6	457.6	8.9	105223	8	AC007399	AC007399 Arabidops
C 7	457.6	8.9	157000	8	AC007627	AC007627 Genomic S
C 8	456.2	8.9	133195	8	AC007123	AC007123 Arabidops
C 9	455.2	8.8	87581	8	ATT7H20	AL162508 Arabidops
C 10	453.4	8.8	158096	8	AC007887	AC007887 Genomic s
C 11	451.8	8.8	108767	8	AC011809	AC011809 Arabidops
C 12	449	8.7	80139	8	AC027135	AC027135 Arabidops
C 13	449	8.7	107200	8	AC006551	AC006551 Arabidops
C 14	449	8.7	110664	8	AC074360	AC074360 Arabidops
C 15	448	8.7	88643	8	AC022522	AC022522 Arabidops
C 16	448	8.7	123386	8	F12F1	AC002131 Arabidops
C 17	446.2	8.7	82652	8	AC007020	AC007020 Arabidops
C 18	445.4	8.7	86139	8	AB028616	AB028616 Arabidops
C 19	442.4	8.6	126386	8	AC005169	AC005169 Arabidops
C 20	441.8	8.6	82451	8	AB073160	AB073160 Arabidops
C 21	439.8	8.5	35308	8	AC079285	AC079285 Arabidops
C 22	439.8	8.5	63773	8	AC074025	AC074025 Arabidops
C 23	436.4	8.5	82053	8	AC022455	AC022455 Arabidops
C 24	435.8	8.5	95581	8	AC015448	AC015448 Arabidops
C 25	433.2	8.4	111767	8	F21E10	AF058914 Arabidops
C 26	430.4	8.4	77636	8	AB077822	AB077822 Arabidops
C 27	430.4	8.4	86014	8	F19C14	AC008051 Sequence
C 28	430.2	8.4	99053	8	AC005292	AC005292 Genomic s
C 29	430	8.4	108387	8	AC016662	AC016662 Arabidops
C 30	429.4	8.3	81580	8	AP000736	AP000736 Arabidops
C 31	429.4	8.3	98017	8	AC027033	AC027033 Arabidops
C 32	429.4	8.3	98948	8	ATF8L21	AL096882 Arabidops
C 33	429.4	8.3	198301	8	ATCHRIV31	AL161531 Arabidops
C 34	428.4	8.3	90284	8	T2K10	AC005966 Arabidops
C 35	427.6	8.3	140680	8	ATF13I12	AL133292 Arabidops
C 36	427.2	8.3	84710	8	AB026643	AB026643 Arabidops
C 37	426.6	8.3	101933	8	AC004146	AC004146 Arabidops
C 38	425.6	8.3	96899	8	AC013288	AC013288 Arabidops
C 39	423.6	8.2	118737	8	T3H13	AF128396 Arabidops
C 40	423.6	8.2	179771	8	ATCHRIV25	AL161513 Arabidops
C 41	422.2	8.2	68889	8	AB010694	AB010694 Arabidops
C 42	421.8	8.2	57246	8	AB010069	AB010069 Arabidops
C 43	419.2	8.1	118737	8	T3H13	AF128396 Arabidops
C 44	419.2	8.1	179771	8	ATCHRIV25	AL161513 Arabidops
C 45	419	8.1	93443	8	ATF2206	AL050300 Arabidops

ALIGNMENTS

RESULT 1  
AF237670  
LOCUS AF237670 5144 bp DNA linear PLN 06-APR-2000  
DEFINITION Arabidopsis thaliana serine/threonine protein kinase SOS2 (SOS2)  
gene, complete cds.  
ACCESSION AF237670  
VERSION AF237670.1 GI:7453644  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 5144)  
AUTHORS Liu,J., Ishitani,M., Halfter,U., Kim,C.S. and Zhu,J.K.





QY	1441	ACAAATTATAAATTGAAGATTCAAAATGCAATGTTAGTTTTTTAAAGTTAATCACCCA	1500
Db	1441	ACAAATTATAAATTGAAGATTCAAAATGCAATGTTAGTTTTTTAAAGTTAATCACCCA	1500
QY	1501	AACGGATAATTGACCCGAACGACTAATTACAGGTGCTATACGGGTACAGTCAAATAACCCG	1560
Db	1501	AACGGATAATTGACCCGAACGACTAATTACAGGTGCTATACGGGTACAGTCAAATAACCCG	1560
QY	1561	ACCCGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTACGGCTTTCGCGA	1620
Db	1561	ACCCGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTACGGCTTTCGCGA	1620
QY	1621	ATCGCATCACGAGCCCTTCCTTCTCTCCTACGCCCTTTTCATCAACCCCTTCTCTGCGAATC	1680
Db	1621	ATCGCATCACGAGCCCTTCCTTCTCTCCTACGCCCTTTTCATCAACCCCTTCTCTGCGAATC	1680
QY	1681	CAATTCTGGTATTTACGATTCTCTTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA	1740
Db	1681	CAATTCTGGTATTTACGATTCTCTTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA	1740
QY	1741	ATTTTATATATGAATATGATCTTGATCCTTACCTTTTTCGTCGTTTCCCTTGATCGAATTT	1800
Db	1741	ATTTTATATATGAATATGATCTTGATCCTTACCTTTTTCGTCGTTTCCCTTGATCGAATTT	1800
QY	1801	ATGTTGATTTATGCTGCGAAATCAAAATTTATGGATACAGAGTAATTTTGTATATATGGAT	1860
Db	1801	ATGTTGATTTATGCTGCGAAATCAAAATTTATGGATACAGAGTAATTTTGTATATATGGAT	1860
QY	1861	GTGTAGGAGCTAAAATTAGCAACAATACGTAAAGTAATTGAAATCGAAATCATAAACGT	1920
Db	1861	GTGTAGGAGCTAAAATTAGCAACAATACGTAAAGTAATTGAAATCGAAATCATAAACGT	1920
QY	1921	TTAAGGAAAGAGGTTTTTTACTAAGTCTCTGAAATAATCTGATTGATAGCTTGTGGTCAATG	1980
Db	1921	TTAAGGAAAGAGGTTTTTTACTAAGTCTCTGAAATAATCTGATTGATAGCTTGTGGTCAATG	1980
QY	1981	GATCAGATAAAGTTTTGTAAAGAAATGACAAAAGAAAATGAGAAGTGGGCAAGTACGAGG	2040
Db	1981	GATCAGATAAAGTTTTGTAAAGAAATGACAAAAGAAAATGAGAAGTGGGCAAGTACGAGG	2040
QY	2041	TTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTTGCGAGGAACACAGACA	2100
Db	2041	TTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTTGCGAGGAACACAGACA	2100
QY	2101	CTGGTGATAATGTAGCCATCAAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAATGG	2160
Db	2101	CTGGTGATAATGTAGCCATCAAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAATGG	2160
QY	2161	TTGATCAGGTAATGTTCTGGATTGTTTTTACATGGAACTAAGGTTGTTGCGTCAATGGT	2220
Db	2161	TTGATCAGGTAATGTTCTGGATTGTTTTTACATGGAACTAAGGTTGTTGCGTCAATGGT	2220
QY	2221	ATGATCTTTGATTTTCGTTTTAAAGCTCTTTACAGATAAAAGAGAGATATCTATAATGAA	2280
Db	2221	ATGATCTTTGATTTTCGTTTTAAAGCTCTTTACAGATAAAAGAGAGATATCTATAATGAA	2280
QY	2281	GATTGTTTCGTACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTCCATGCA	2340
Db	2281	GATTGTTTCGTACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTCCATGCA	2340
QY	2341	TCTGCGAAATTTTATCTCTGAAGTGTGTTTTCGATCATGTTCTCTGTTGTTTTTTTGTG	2400
Db	2341	TCTGCGAAATTTTATCTCTGAAGTGTGTTTTCGATCATGTTCTCTGTTGTTTTTTTGTG	2400
QY	2401	ATTTTCCCGATGTAGGTGTTGGCGAGTCTTCGAAAATATATATAGTTTTGGAGTTTGTG	2460
Db	2401	ATTTTCCCGATGTAGGTGTTGGCGAGTCTTCGAAAATATATATAGTTTTGGAGTTTGTG	2460
QY	2461	ACAGGAGGAGACTCTTTTGATAGAAATGTACGGAACTTCCATACTTGTAGGCAGCGTCCA	2520
Db	2461	ACAGGAGGAGACTCTTTTGATAGAAATGTACGGAACTTCCATACTTGTAGGCAGCGTCCA	2520
QY	2521	TTAGTTAAACCTCTCTACTTAAATTTTTTAAATATATGAAATCTTTTCATGCAGGTTCATAA	2580

Db	2521	TTAGTTAAACCTCTCTACTTAAATTTTTTAAATATATGAAATCTTTTCATGCAGGTTCATAA	2580
QY	2581	AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATGCTGTTGC	2640
Db	2581	AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATGCTGTTGC	2640
QY	2641	TCATTTGTCACTGCAAGGGTGTTTTACCACCGTGACCTAAAGGTAAGACGTTGTTTGTGTTT	2700
Db	2641	TCATTTGTCACTGCAAGGGTGTTTTACCACCGTGACCTAAAGGTAAGACGTTGTTTGTGTTT	2700
QY	2701	ACCAATATTCCTCAGAAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTGTGTCGC	2760
Db	2701	ACCAATATTCCTCAGAAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTGTGTCGC	2760
QY	2761	TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATFGGAAATCTGAAGGTTT	2820
Db	2761	TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATFGGAAATCTGAAGGTTT	2820
QY	2821	CGGATTTCCGACTCAGTGCAATGCTCAGGAAAGTAAGTGCTCTTATCTCTGCTTCAGCAG	2880
Db	2821	CGGATTTCCGACTCAGTGCAATGCTCAGGAAAGTAAGTGCTCTTATCTCTGCTTCAGCAG	2880
QY	2881	TCTGCTTACGTGGTCATTAACTTGTATATATACTCAATCAGGGAGTAGAACTTCTGCGTAC	2940
Db	2881	TCTGCTTACGTGGTCATTAACTTGTATATATACTCAATCAGGGAGTAGAACTTCTGCGTAC	2940
QY	2941	CACATGTGGAACCTCCGAACCTATGTAGCTCCAGAGGTACTTAGTGCACAGGGTTACGATGG	3000
Db	2941	CACATGTGGAACCTCCGAACCTATGTAGCTCCAGAGGTACTTAGTGCACAGGGTTACGATGG	3000
QY	3001	TTCAGCAGCTGATATTTTGGTCTTTCGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT	3060
Db	3001	TTCAGCAGCTGATATTTTGGTCTTTCGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT	3060
QY	3061	ACCTTTTTCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTTCGG	3120
Db	3061	ACCTTTTTCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTTCGG	3120
QY	3121	GAAGAAATCATGAATTCCTTGTCTATGCTTTTGTCAAACCGTTTATGATTGGTTTTTGC	3180
Db	3121	GAAGAAATCATGAATTCCTTGTCTATGCTTTTGTCAAACCGTTTATGATTGGTTTTTGC	3180
QY	3181	AATTTCAACCTTAGTTTTTGAGCTTTTACACATTTGTTATTTTACAGATAAATGCAGCAGT	3240
Db	3181	AATTTCAACCTTAGTTTTTGAGCTTTTACACATTTGTTATTTTACAGATAAATGCAGCAGT	3240
QY	3241	TTTTCTTGTCCACCGTGGTTTTCCGCAGAAAGTGAAAGTTTTTAAATACATAGGATACCTTGACC	3300
Db	3241	TTTTCTTGTCCACCGTGGTTTTCCGCAGAAAGTGAAAGTTTTTAAATACATAGGATACCTTGACC	3300
QY	3301	CCAAATCCCAAAACAGTGAGTATTTTGTCTCTCTAGCTATCAGGTTTTTGGTGAT	3360
Db	3301	CCAAATCCCAAAACAGTGAGTATTTTGTCTCTCTAGCTATCAGGTTTTTGGTGAT	3360
QY	3361	ATTTAATGTTCTAGTAATTATATCTGTTTATCTAATTATGTTTCTCAATTAGAGCAGCGT	3420
Db	3361	ATTTAATGTTCTAGTAATTATATCTGTTTATCTAATTATGTTTCTCAATTAGAGCAGCGT	3420
QY	3421	ATTCAAATTCAGGAAATCAAGAAAGATCCTTGGTTCAGATTAAATTTATGTCCCTATACGA	3480
Db	3421	ATTCAAATTCAGGAAATCAAGAAAGATCCTTGGTTCAGATTAAATTTATGTCCCTATACGA	3480
QY	3481	GCAAGGGAAGAAGAAGACTGAATTTGGATGATATTTGTCGAGTTTTTGTGATGGAATTGAG	3540
Db	3481	GCAAGGGAAGAAGAAGACTGAATTTGGATGATATTTGTCGAGTTTTTGTGATGGAATTGAG	3540
QY	3541	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCATTAAATGTTAAGAT	3600
Db	3541	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCATTAAATGTTAAGAT	3600
QY	3601	CTCTTAGGAACGTTTGGATGACCAAGAAAGAGTGTGCTACAGGATAGAACAAAAATAGT	3660

Db 3601 CTCCTTAGGAACGTTTGGATGACCAAGAAGAGTGTGTTGCTACAGGATAGAACAAATAGT 3660

Qy 3661 AGGCATGTGTTAGTTACCAAACTGTAAACTGCTTCTTTATTCAATTCGCCAAACCATAG 3720

Db 3661 AGGCATGTGTTAGTTACCAAACTGTAAACTGCTTCTTTATTCAATTCGCCAAACCATAG 3720

Qy 3721 ACCTTAGGAAGACTTAGATTACAAGAGAAATCTCTATTCTCGACCAAAACCCCTAGACA 3780

Db 3721 ACCTTAGGAAGACTTAGATTACAAGAGAAATCTCTATTCTCGACCAAAACCCCTAGACA 3780

Qy 3781 AAATCCAGAATACCCCTAGGGCTAAATTACAATGTTCAATGTTCAATGTTCAATGTTCAAT 3840

Db 3781 AAATCCAGAATACCCCTAGGGCTAAATTACAATGTTCAATGTTCAATGTTCAATGTTCAAT 3840

Qy 3841 TGCTAAGATTGTTCTTAGCTTTGGTATAGCCTAGATATATATAGATACCGTAAATTTCTAAT 3900

Db 3841 TGCTAAGATTGTTCTTAGCTTTGGTATAGCCTAGATATATATAGATACCGTAAATTTCTAAT 3900

Qy 3901 GCATATTTAGAAATGTTTACTTAAACTCAGGCTCTGCTCTTCTAAACTGTTACTTCAA 3960

Db 3901 GCATATTTAGAAATGTTTACTTAAACTCAGGCTCTGCTCTTCTAAACTGTTACTTCAA 3960

Qy 3961 TTGTTAAACTAAACCTCAGTATCTGTCTTAGCTAAAGTTACTTTTACTTTGTTTTCATT 4020

Db 3961 TTGTTAAACTAAACCTCAGTATCTGTCTTAGCTAAAGTTACTTTTACTTTGTTTTCATT 4020

Qy 4021 AAGTTGACCTGTCAATTGCACTTGTTCACAGGCGAGTATGTAGCGGAGATGTAGAGAG 4080

Db 4021 AAGTTGACCTGTCAATTGCACTTGTTCACAGGCGAGTATGTAGCGGAGATGTAGAGAG 4080

Qy 4081 AAATGATGAAGGGCCCTGATGATGAATGCTTTTGAGATGATTACCTTATCACAAGGCTT 4140

Db 4081 AAATGATGAAGGGCCCTGATGATGAATGCTTTTGAGATGATTACCTTATCACAAGGCTT 4140

Qy 4141 AAATTTATCTGCACTATTTGACAGCGCAGGAGTAGTACCTGATTTTCTATTACTGGTCA 4200

Db 4141 AAATTTATCTGCACTATTTGACAGCGCAGGAGTAGTACCTGATTTTCTATTACTGGTCA 4200

Qy 4201 AGAGATCTCCATTTCGAATAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCGTTT 4260

Db 4201 AGAGATCTCCATTTCGAATAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCGTTT 4260

Qy 4261 TGACTGCCCTTATGATGCTGTGTTCTTAGTTTGTATATAAATACTATAAGTTTCATTAGTA 4320

Db 4261 TGACTGCCCTTATGATGCTGTGTTCTTAGTTTGTATATAAATACTATAAGTTTCATTAGTA 4320

Qy 4321 TTGTTTGCATGGCATTAGTAGATACAAATGGAATCCAAATGTTTCTGTCATATTGATGGC 4380

Db 4321 TTGTTTGCATGGCATTAGTAGATACAAATGGAATCCAAATGTTTCTGTCATATTGATGGC 4380

Qy 4381 TGATCCCTTTGATCTCGCAGGATTTTGTAAAGGCAACCCGTTTGTCTCGAAGGGA 4440

Db 4381 TGATCCCTTTGATCTCGCAGGATTTTGTAAAGGCAACCCGTTTGTCTCGAAGGGA 4440

Qy 4441 ACCTAGTGAGATAAATTGCTAAACATGAGGCTGTAGCGAATCAATGGGTTTAAAGTCTCA 4500

Db 4441 ACCTAGTGAGATAAATTGCTAAACATGAGGCTGTAGCGAATCAATGGGTTTAAAGTCTCA 4500

Qy 4501 TACACGAACTTCAAGGTAACGAATTCCTAGCATATTACACTATCAGAGAGATTATGCA 4560

Db 4501 TACACGAACTTCAAGGTAACGAATTCCTAGCATATTACACTATCAGAGAGATTATGCA 4560

Qy 4561 TTATTTTAAACTCTCAACTGTTTAAACGATGTGTAGATAGATTGATAAGATTGACAAGG 4620

Db 4561 TTATTTTAAACTCTCAACTGTTTAAACGATGTGTAGATAGATTGATAAGATTGACAAGG 4620

Qy 4621 AAACCTTAGTTTATCTCTGGCGTTTCAAAAACGAAGTCTAGTGTGAATTATCATTTT 4680

Db 4621 AAACCTTAGTTTATCTCTGGCGTTTCAAAAACGAAGTCTAGTGTGAATTATCATTTT 4680

Qy 4681 AATGTTAGCAGAGAGTACAATTGTTTATGATTTGTTACGTCATGCTCAACAGACAGG 4740

Db 4681 AATGTTAGCAGAGAGTACAATTGTTTATGATTTGTTACGTCATGCTCAACAGACAGG 4740

Qy 4741 CTCGAGGATTTATCTTCGATCAAGGCCGACAGTTAGCTGTTGTGATAGAGTAATTATT 4800

Db 4741 CTCGAGGATTTATCTTCGATCAAGGCCGACAGTTAGCTGTTGTGATAGAGTAATTATT 4800

Qy 4801 GCTTGTGTGATTGTAATATAAGTTTGTCTTTCAGTTTAAGGGTATCTAGCAAATTG 4860

Db 4801 GCTTGTGTGATTGTAATATAAGTTTGTCTTTCAGTTTAAGGGTATCTAGCAAATTG 4860

Qy 4861 AAATTAACCTACATGCAGATTTACGAGGTGGCACCATTGCTTTTTCATGTTAGCGTAAGA 4920

Db 4861 AAATTAACCTACATGCAGATTTACGAGGTGGCACCATTGCTTTTTCATGTTAGCGTAAGA 4920

Qy 4921 AAGGCTGCTGTTGAAACTCTTGAATATCACAAAGTTTATAATATATATATATATATAT 4980

Db 4921 AAGGCTGCTGTTGAAACTCTTGAATATCACAAAGTTTATAATATATATATATATATAT 4980

Qy 4981 AGTTGCATCATTTACTGTGTTGCGGATTAGAGTGAATTTTGTGTTGTTGTTGTTGTTGTT 5040

Db 4981 AGTTGCATCATTTACTGTGTTGCGGATTAGAGTGAATTTTGTGTTGTTGTTGTTGTTGTT 5040

Qy 5041 TACAAGAAGCTATGTTTCGAAACTGGAAACATATATATGAGGGCAACAGAGGAATACCA 5100

Db 5041 TACAAGAAGCTATGTTTCGAAACTGGAAACATATATATGAGGGCAACAGAGGAATACCA 5100

Qy 5101 AAGTCAGAGATTCTCAGAACAAATCACGTTTTCATCCCAACTTAA 5144

Db 5101 AAGTCAGAGATTCTCAGAACAAATCACGTTTTCATCCCAACTTAA 5144

RESULT 2

AB025611 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21B8. 50015 bp DNA linear

DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K21B8.

ACCESSION AB025611 BA000015

VERSION AB025611.1 GI:4589417

KEYWORDS

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

1 (sites) Kaneko, T., Katoh, T., Asamizu, E., Sato, S., Nakamura, Y., Kotani, H. and Tabata, S.

Structural analysis of Arabidopsis thaliana chromosome 5. XI

AUTHORS

Unpublished

2 (bases 1 to 50015)

REFERENCE

Nakamura, Y.

AUTHORS

Direct Submission

TITLE

Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)

JOURNAL

Address for correspondence: kaos@kazusa.or.jp

For the latest information on annotation of this clone, please see [http://www.kazusa.or.jp/kaos/cgi-bin/agd\\_graph.cgi?c=K21B8](http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=K21B8)

Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.

The software programs used to predict genes include: Grail

(Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/Grail-1.3/>),

GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and

SplicePredictor (Volker Brendel, Stanford University, <http://gremmlin.zool.iastate.edu/cgi-bin/sp.cgi>).

Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAscan-SE/>).

This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.

FEATURES	The 5' clone is F6I13 and the 3' clone is MOK9.	
	Location/Qualifiers	
source	1..50015	
	/organism="Arabidopsis thaliana"	
CDS	/mol_type="genomic DNA"	
	/strain="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="5"	
	/clone="K21B8"	
	/clone_lib="Mitsui TAC"	
	2444..2620	
	/note="gene_id:K21B8.1	
	unknown protein"	
	/pseudo	
	/codon_start=1	
	/evidence=not_experimental	
CDS	join(21353..21804,21890..21991)	
	/note="gene_id:K21B8.2	
	pir  T08458	
	similar to unknown protein"	
	/pseudo	
	/codon_start=1	
	/evidence=not_experimental	
	join(23764..23928,24015..24077,24176..24247,24333..24440,	
	24538..24612,24681..24860,24985..25074,25178..25300,	
	25812..25931,26160..26276,26495..26551,26639..26713,	
	26798..26893)	
	/note="gb AAF62923.1	
	gene_id:K21B8.3"	
	/codon_start=1	
	/evidence=not_experimental	
	/product="serine/threonine protein kinase SOS2"	
	/protein_id="BAA98146.1"	
	/db_xref="GI:8978255"	
	/translation="MTKKMRVRVGKYEVRTIGEGTFAKVKFPARNTDGTGDNVAIKINAK	
	STILKNRMVDQIKREISIMKIIVHPNIVRLYEVLASPSKIYIVLFFVTGGELFDRIVH	
	KGRLESESRKYFQQLVDAVAHCHKGVYHRLDKPENLLDITGNLKVSDFGLSALPQ	
	EGVELLRTTCGTPNYVAPEVLSQGYDGSADIWSCGVILFVILAGYLPFSETDLPGL	
	YRKINAAEFSCPPWFSAEVKFLIHLIDPNPKTRIQIQGIKKDPWFRNLNVPIRAREE	
	EVNLDLDIRAVFDGIEGSYVAENVERNDEGPLMNAFEMITLSQGLNLSALFDRRQDF	
	VKROTFRVSRREPSEIIANIEAVANSMGFKSHTRNFKRLEGLSSI KAGQLAVVIEIY	
	EVAPSLFMVDVRKAAGETLEHYKFKKLCCKLENIIWRATEGIPKSEILRTIIF"	
	complement(28006..33201)	
	/note="contains similarity to Tall-like non-LTR	
	retroelement protein	
	gene_id:K21B8.4"	
	/pseudo	
	/codon_start=1	
	/evidence=not_experimental	
	complement(join(33610..34901,35547..37507,37633..38034))	
	/note="gene_id:K21B8.5"	
	/pseudo	
	/codon_start=1	
	/evidence=not_experimental	
	complement(join(38377..38751,38917..39776))	
	/note="gene_id:K21B8.6"	
	/pseudo	
	/codon_start=1	
	/evidence=not_experimental	
	/product="Tall-like non-LTR retroelement protein-like"	
	complement(join(44026..44296,44491..44768))	
	/note="gb AAD56332.1	
	gene_id:K21B8.7	
	similar to unknown protein"	
	/codon_start=1	
	/evidence=not_experimental	
	/protein_id="BAA98147.1"	
	/db_xref="GI:8978256"	
	/translation="MDPIKLSERLFELGTIEGTKKVNAYTVFHWIKDKPVLIDDHLK	
	KLENSQFGSILQMVDDLNSVRVMHHILTRQLVTKKEFELWWT FAGKPMRLALLVIVEG	
	ILCPTIITYILPQVVNMHSDLDEFIKFPWGRESFLLTVRSVKTRKLYTNSQGSIAI	

QGFHAITLIFRIWVPSKRLV"	
complement(join(48713..49236,49497..49565))	
/note="gb AAD15517.1	
gene_id:K21B8.8	
similar to unknown protein"	
/pseudo	
/codon_start=1	
/evidence=not_experimental	
BASE COUNT	16169 a 8857 c 7700 g 17289 t
ORIGIN	
Query Match 100.0%; Score 5144; DB 8; Length 50015;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 5144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 TTTTTCGCGGAAAAATCTCGGGTTTACGTTTTCGCGGGGAAAAATCTCGTGTTTACGTTTT 60
Db	21761 TTTTTCGCGGAAAAATCTCGGGTTTACGTTTTCGCGGGGAAAAATCTCGTGTTTACGTTTT 21820
QY	61 TGGCGGAAAAATCTCGGGTTTACGTTTTCGCGAGAAAAATCACGGGTTTACGTTTTTG 120
Db	21821 TGGCGGAAAAATCTCGGGTTTACGTTTTCGCGAGAAAAATCACGGGTTTACGTTTTTG 21880
QY	121 CGGAAAAATCACGGATTATGTTTTTGGTGGAAAAAATACGAGTTTACGTTTCTCAAT 180
Db	21881 CGGAAAAATCACGGATTATGTTTTTGGTGGAAAAAATACGAGTTTACGTTTCTCAAT 21940
QY	181 TTCAATGCTTGATATTTAAGAAATTTGAAAAATATTAAATTTTAAATGGTTTGA 240
Db	21941 TTCAATGCTTGATATTTAAGAAATTTGAAAAATATTAAATTTTAAATGGTTTGA 22000
QY	241 TGTGTTGGTTAAACCTAAATTCGCATTTGGTTTAGAGATTTTAGTGGTTTATTCAATTT 300
Db	22001 TGTGTTGGTTAAACCTAAATTCGCATTTGGTTTAGAGATTTTAGTGGTTTATTCAATTT 22060
QY	301 TACAAAAATTTGATGGGTTAATTCGATAAACCATGGAAACCATTAACCATTACAACCTAAC 360
Db	22061 TACAAAAATTTGATGGGTTAATTCGATAAACCATGGAAACCATTAACCATTACAACCTAAC 22120
QY	361 TCATTTTACTCATCAAAACCAATGACTCATCAACTCATTTGACTCATCAACTCATTTGAG 420
Db	22121 TCATTTTACTCATCAAAACCAATGACTCATCAACTCATTTGACTCATCAACTCATTTGAG 22180
QY	421 TCAAAAATTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480
Db	22181 TCAAAAATTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 22240
QY	481 TTGACCCATTTTGACACCCCTACATATGATCATATAAGTTTAATAATCAAAATTACTATTGA 540
Db	22241 TTGACCCATTTTGACACCCCTACATATGATCATATAAGTTTAATAATCAAAATTACTATTGA 22300
QY	541 TAACTTTTTAACGGAATTGTTTGTAAAGTTTCATTTACGTTATTTATATAAAAAACATA 600
Db	22301 TAACTTTTTAACGGAATTGTTTGTAAAGTTTCATTTACGTTATTTATATAAAAAACATA 22360
QY	601 ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGCAAAATATTACTGTAA 660
Db	22361 ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGCAAAATATTACTGTAA 22420
QY	661 TACTTTTTCACTTTATCAATATTTTCTTATATTCATTTAGCCACCAACCTACATTT 720
Db	22421 TACTTTTTCACTTTATCAATATTTTCTTATATTCATTTAGCCACCAACCTACATTT 22480
QY	721 TCATATTTCTCTTCACATTCATTTATTTCTTGCTCTTGATTCCTTCTTATCATCTTCAT 780
Db	22481 TCATATTTCTCTTCACATTCATTTATTTCTTGCTCTTGATTCCTTCTTATCATCTTCAT 22540
QY	781 TGCCAATTTTTTCATTGACATTGTCATCGTTTACTTTTCAATTTAGCCACCAACCTACATTT 840
Db	22541 TGCCAATTTTTTCATTGACATTGTCATCGTTTACTTTTCAATTTAGCCACCAACCTACATTT 22600
QY	841 TGTGAATCCCAACCGACTATAGGACTATCAAGAGTTTTCAAACTTTTAAAAAAGATCT 900



Db 22601 TGTGAATCCCAACCGACTATAGGACTATCAAGAAGTTTTTCAAACTTTTTAAAAAAGATCT 22660

Qy 901 TGAACCTTTAAAGCAAATCCCAACCGACTATAGAACCATAGTCTCTTCTCTTGTGATGA 960

Db 22661 TGAACCTTTAAAGCAAATCCCAACCGACTATAGAACCATAGTCTCTTCTTGTGATGA 22720

Qy 961 AGCTCTTCTCGTCCGGCGAAAAATCTAGCCATAAAAGCCTCTTCAACATCACCTAGTAT 1020

Db 22721 AGCTCTTCTCGTCCGGCGAAAAATCTAGCCATAAAAGCCTCTTCAACATCACCTAGTAT 22780

Qy 1021 ATTGACCGTGACCATCTTTTGGACCAATGCTTTGTGAATGAACCGTCGATAAACCGTGT 1080

Db 22781 ATTGACCGTGACCATCTTTTGGACCAATGCTTTGTGAATGAACCGTCGATAAACCGTGT 22840

Qy 1081 ATCACTTACGCCAAATTTTCCCTAGTGTGTTTCCAAACTCTCACGAATCCTTATCGA 1140

Db 22841 ATCACTTACGCCAAATTTTCCCTAGTGTGTTTCCAAACTCTCACGAATCCTTATCGA 22900

Qy 1141 ACTTTTTATATATACACTTTGTAGCAATGGAAAAGTATCTTTGTATGCTTTGTCTTAAACT 1200

Db 22901 ACTTTTTATATATACACTTTGTAGCAATGGAAAAGTATCTTTGTATGCTTTGTCTTAAACT 22960

Qy 1201 TAGACATCCTTGTTCTCTTGGTTTTTTTGAACCTTGCTTGACTTAAATGAAGTTAAAAAT 1260

Db 22961 TAGACATCCTTGTTCTCTTGGTTTTTTTGAACCTTGCTTGACTTAAATGAAGTTAAAAAT 23020

Qy 1261 TTGTAGTTAAAAATAGAAAAATTTTACTAAATTTGAGTTGATTCGATTAATCATAGCTAGATAAT 1320

Db 23021 TTGTAGTTAAAAATAGAAAAATTTTACTAAATTTGAGTTGATTCGATTAATCATAGCTAGATAAT 23080

Qy 1321 TTGAAAAAAATTAATAATAAATTTTGAAAAATACTATATGTTTATTTTTTAAAAATAATTTA 1380

Db 23081 TTGAAAAAAATTAATAATAAATTTTGAAAAATACTATATGTTTATTTTTTAAAAATAATTTA 23140

Qy 1381 CTAATTTATTAATTAAGTATATTAAGTTAGTTAGTTTTTTTTTCTTTTTTAAATTTTGAAAAACCTC 1440

Db 23141 CTAATTTATTAATTAAGTATATTAAGTTAGTTAGTTTTTTTTTCTTTTTTAAATTTTGAAAAACCTC 23200

Qy 1441 ACAATTTATTAATTAAGATTCAAAATGCAATGTTAGTTTTTTTAAAGTTTAAATCACCCCA 1500

Db 23201 ACAATTTATTAATTAAGATTCAAAATGCAATGTTAGTTTTTTTAAAGTTTAAATCACCCCA 23260

Qy 1501 AACGGATAATTGACCCGAACGACTTAATTCAGGTCGTATACGGGTACAGTCAAATAACCCG 1560

Db 23261 AACGGATAATTGACCCGAACGACTTAATTCAGGTCGTATACGGGTACAGTCAAATAACCCG 23320

Qy 1561 ACCCGAATGCTCAACGGATCTGCACGACGTTTCGCATCGACGATTCACGGCTTTCGCGA 1620

Db 23321 ACCCGAATGCTCAACGGATCTGCACGACGTTTCGCATCGACGATTCACGGCTTTCGCGA 23380

Qy 1621 ATCGCATACGAGCCTTCCTTCTCTCCACGCCCTCTTTCATCAACCCCTCTCTGCGAATC 1680

Db 23381 ATCGCATACGAGCCTTCCTTCTCTCCACGCCCTCTTTCATCAACCCCTCTCTGCGAATC 23440

Qy 1681 CAATTCGGTATTTACAGATTCTCTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA 1740

Db 23441 CAATTCGGTATTTACAGATTCTCTTTCGACTACGCCCAATCGCAATTCAGCCGTACGA 23500

Qy 1741 ATTTTATATATTAATGATTAATGATTCCTTACCTTTTCGTCGTTTCTTGATCGAATTT 1800

Db 23501 ATTTTATATATTAATGATTAATGATTCCTTACCTTTTCGTCGTTTCTTGATCGAATTT 23560

Qy 1801 ATGTTGATTTATGCTCGGAAATCAAAATTAATGGATACAGAGTAATTTTGATATATGGAT 1860

Db 23561 ATGTTGATTTATGCTCGGAAATCAAAATTAATGGATACAGAGTAATTTTGATATATGGAT 23620

Qy 1861 GTGTAGGAGCTAAAAATTAGCAAACAATACGTAAAGTAATTGAATCGAAATCATAAACGT 1920

Db 23621 GTGTAGGAGCTAAAAATTAGCAAACAATACGTAAAGTAATTGAATCGAAATCATAAACGT 23680

Qy 1921 TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTGATAGCTTGTGGTCAATG 1980

Db 23681 TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTGATAGCTTGTGGTCAATG 23740

Qy 1981 GATCAGATAAAAAGTTTGTAAAGAATGACAAAGAAAAATGAGAAGAGTGGCAAGTACGAGG 2040

Db 23741 GATCAGATAAAAAGTTTGTAAAGAATGACAAAGAAAAATGAGAAGAGTGGCAAGTACGAGG 23800

Qy 2041 TTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAAGTTTAAAGTTTCGAGGAACACAGACA 2100

Db 23801 TTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAAGTTTAAAGTTTCGAGGAACACAGACA 23860

Qy 2101 CTGGTGATAATGTAGCCATCAAAATTAATGGCTAAGAGTACAAATCTTAAGAACAGAAATGG 2160

Db 23861 CTGGTGATAATGTAGCCATCAAAATTAATGGCTAAGAGTACAAATCTTAAGAACAGAAATGG 23920

Qy 2161 TTGATCAGGTATGTTCTGGATGTTTTTTTACATGGAAACTAAAGTTGTTGCGTCAATGGT 2220

Db 23921 TTGATCAGGTATGTTCTGGATGTTTTTTTACATGGAAACTAAAGTTGTTGCGTCAATGGT 23980

Qy 2221 ATGATCTTTTGATTTTCGTTTAAAGCTCTTTTACAGATAAAAAGAGAGATATCTATAATGAA 2280

Db 23981 ATGATCTTTTGATTTTCGTTTAAAGCTCTTTTACAGATAAAAAGAGAGATATCTATAATGAA 24040

Qy 2281 GATTGTTGTCACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTCCATGCA 2340

Db 24041 GATTGTTGTCACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTCCATGCA 24100

Qy 2341 TCTGCGAAAAATTTATCTCTGAAGTGTTTTGTGCATCATTTGTCATCATTTGTTGTTTTTTGTG 2400

Db 24101 TCTGCGAAAAATTTATCTCTGAAGTGTTTTGTGCATCATTTGTCATCATTTGTTGTTTTTTGTG 24160

Qy 2401 ATTTTCCCGATGTAGGTGTTGCGGAGTCCTTCGAAAAATATATATAGTTTTTGGAGTTTTGTG 2460

Db 24161 ATTTTCCCGATGTAGGTGTTGCGGAGTCCTTCGAAAAATATATATAGTTTTTGGAGTTTTGTG 24220

Qy 2461 ACAGGAGGAGAGCTCTTTTGATAGAAATTGTACGGAACTTCCATACATTGTAGGCAGGTCCTA 2520

Db 24221 ACAGGAGGAGAGCTCTTTTGATAGAAATTGTACGGAACTTCCATACATTGTAGGCAGGTCCTA 24280

Qy 2521 TTAGTTAAAAACCTCTCTACTTAAATTTTAAATATATGAATCTTTTCATGCAGGTTTCATAA 2580

Db 24281 TTAGTTAAAAACCTCTCTACTTAAATTTTAAATATATGAATCTTTTCATGCAGGTTTCATAA 24340

Qy 2581 AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATGCTGTTGC 2640

Db 24341 AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACTTTCAACAGCTTGTAGATGCTGTTGC 24400

Qy 2641 TCATTTGTCACGTCAAGGGTGTTTACCACCGTGACCTAAAGSTAAAAGACGTGTTTTTGTGTT 2700

Db 24401 TCATTTGTCACGTCAAGGGTGTTTACCACCGTGACCTAAAGSTAAAAGACGTGTTTTTGTGTT 24460

Qy 2701 ACCAATATTTCTCAGAAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTTTGTGTCGC 2760

Db 24461 ACCAATATTTCTCAGAAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTTTGTGTCGC 24520

Qy 2761 TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAATGGAAATCTGAAGGTTT 2820

Db 24521 TATGTTATGTTATCTAGCCAGAAAAATCTTTTACTCGATACAAATGGAAATCTGAAGGTTT 24580

Qy 2821 CGGATTTTCGGACTCAGTGCATTCGCTCAGGAAAGTAAGTCTCTTATCTCTGCTTCAGCAG 2880

Db 24581 CGGATTTTCGGACTCAGTGCATTCGCTCAGGAAAGTAAGTCTCTTATCTCTGCTTCAGCAG 24640

Qy 2881 TCTGCTTACGTGGTCAATTAATGTTTATATATCTCAATCAGGGAGTAGAACTTCTGCGTAC 2940

Db 24641 TCTGCTTACGTGGTCAATTAATGTTTATATATCTCAATCAGGGAGTAGAACTTCTGCGTAC 24700

Qy 2941 CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTAATTTAGTGGACAGGGTTACGATGG 3000

Db 24701 CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTAATTTAGTGGACAGGGTTACGATGG 24760

Qy 3001 TTCAGCAGCTGATATTTTGGTCTTTCGGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT 3060

Db 24761 TTCAGCAGCTGATATTTTGGTCTTTCGGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT 24820



QY	3061	ACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTCCG	3120
Db	24821		
QY	3121	GAAGAAATCATGAATTCCTTGTTCATGGCTTTTGTCAAAACCGTTTATTGATTTGGTTTGC	3180
Db	24881		
QY	3181	AATTTCAACCTTAGTTTGAGCTTTTACACATTGTTATTTACAGATAAAATGCAGCAGAGT	3240
Db	24941		
QY	3241	TTTCTTGTCCACCGTGGTTTCCGCGAGAAGTGAAGTTTAAATACATAGGATACTTGACC	3300
Db	25001		
QY	3301	CCAAATCCCAAAACAGTGAGTATTTTGCTTTGTCTCTCCTAGCTATCAGGTTTGGTGAT	3360
Db	25061		
QY	3361	ATTAAATGTTCTAGTAATATATCTGTTTATCTAATATTTGTTTCTCAATTAGAGCAGCGT	3420
Db	25121		
QY	3421	ATTCAAATCAAGGAATCAAGAAAGATCCTTGGTTTCAGATTAAATTATGTGCCTATACGA	3480
Db	25181		
QY	3481	GCAAGGGAAGAAGAAAGTGAATTTGGATGATATTCTGTCAGTTTTTTCATGGAATTGAG	3540
Db	25241		
QY	3541	GTTTGTGCTTCGCTTCATTATATATGCTCTTTGCTGGTCAATTCATTTAAATGTTAAGAT	3600
Db	25301		
QY	3601	CTCTTAGGAACGTTTGGATGACCAAGAAGATGTTTGCTACAGGATAGAACAATAAGT	3660
Db	25361		
QY	3661	AGGCATGTGTTAGTTACCAACCTGTAACTGCTCTTTATTCAATTCGCCAAACCATAG	3720
Db	25421		
QY	3721	ACCTTAGGAAGACTTAGATTTACAAGAGAAATTCCTATTCTCGACCAAAAACCCCTAGACA	3780
Db	25481		
QY	3781	AAATCCAGAATACCCCTAGGGCTAATTACAATGTTTCATGTACCTATCAATATATATCTCC	3840
Db	25541		
QY	3841	TGCTAAGATTGCTTAGCTTTGGTATAGCCTAGATATATATAGATACCGTAATTTCTAAT	3900
Db	25601		
QY	3901	GCATATTTAGAATGTTTTTACTTAAACTCAGGCTCTTGCTCTTCTAATAACTTGTACTTCAA	3960
Db	25661		
QY	3961	TTGTTAAACTPAAAAACCTCAGTATCTGTCTTAGCTAAAGTTACTTTTACTTGTTTTCAAT	4020
Db	25721		
QY	4021	AAGTTGACCTGTCAATTGCACCTTGTTCACAGGGCAGTTATGTAGCGGAGAAATGTAGAGAG	4080
Db	25781		
QY	4081	AAATGATGAAGGGCCCTGTATGATGAATGCCTTTTGATGATGATTAACCTTATCACAAGGCTT	4140
Db	25841		
QY	4141	AAATTTATCGCACTATTTTGACAGGCGACAGGTAGTACCTGAATTTTCTATTACTGGTCAAT	4200

Db	25901		
QY	4201	AGAGATCTCCATTTCCGAATAAAAGAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT	4260
Db	25961		
QY	4261	TGACTGCCCTTATCATGCTGTGTTCTTAGTTTGTATAATAACTATAAGTTCATTAGATGA	4320
Db	26021		
QY	4321	TTGGTTGCATGGCATTAGTAGATACAAATGGAATCCAAAATGTTCCCTGCATATTGATGGC	4380
Db	26081		
QY	4381	TGATCCTTTGATCTCGCAGGATTTTGTAAAAAGGCAAAACCCGTTTGTGTTCTCGAAGGGA	4440
Db	26141		
QY	4441	ACCTAGTGAGATAAATGCTAAACATTTAGGCTGTAGCGAACTCAATGGGTTTAAAGTCTCA	4500
Db	26201		
QY	4501	TACACGAAACTTCAAGGTAAACGAATTCCTAGCATATTACACTTATCACAGAGATTATGCA	4560
Db	26261		
QY	4561	TTATTTTAAAACTCTCAACTGTTTAAACGCATGTGTAGATAGATTGATAAGATTGACAAGG	4620
Db	26321		
QY	4621	AAACTTAGTTTATATCTCTGGCGTTCAAAAACGAAAAGTCCCTAGTGTGAATTTATCATTTT	4680
Db	26381		
QY	4681	AATGTTAGCAGAGAGTACAATTGTTATGATTTGTTACGTCATATGCTCAACAGACAAGG	4740
Db	26441		
QY	4741	CTCAGGGATTATCTTCGATCAAGCCGGACAGTTAGCTGTTGTGATAGAGGTAATTATT	4800
Db	26501		
QY	4801	GCITTGTTGTGATTGTAATATAAGTTTGCTTTGCTTCAGTTTAAAGGTATCTAGCAAAATG	4860
Db	26561		
QY	4861	AAATTAACTTACATGCAGATTTACGAGGTGGCACCATCGCTTTTCATGGTAGACGTAAGA	4920
Db	26621		
QY	4921	AAGCTGCTGGTGAAACTCTTGAATATCACAAAGGTTTATAAATATATATCCAAATAACAT	4980
Db	26681		
QY	4981	AGTTGCATCATTAAGTTGTTGCGGATTAGAGTGATATTGTTGTTTGTGGTATCGCAGTTC	5040
Db	26741		
QY	5041	TACAAGAAAGCTATGTTTCGAAACTGGAACAATAATATGAGGGGCAACAGAAAGGAATACCA	5100
Db	26801		
QY	5101	AAGTCAGAGATTCTCAGAAACAATCACGTTTTTGTATCCCAACTTAA	5144
Db	26861		

RESULT 3  
F6113  
LOCUS  
DEFINITION Arabidopsis thaliana BAC F6113.  
ACCESSION AF262044

linear  
16228 bp  
DNA  
PLN 26-JUL-2000



QY	961	AGCTCTTCTCGTGCCGCGGAAAAATCTAGGCCATAAAAGCCTCTTCAACATCACCTAGTAT	1020
Db	12662	AGCTCTTCTCGTGCCGCGGAAAAATCTAGGCCATAAAAGCCTCTTCAACATCACCTAGTAT	12721
QY	1021	ATTGACCGTGACCATCTTTTGTGACCATTGCCTTTGTGAATGAACCGTCGATAAACCGTGTT	1080
Db	12722	ATTGACCGTGACCATCTTTTGTGACCATTGCCTTTGTGAATGAACCGTCGATAAACCGTGTT	12781
QY	1081	ATCACTTACGCCAAATTTTCCCTAGTGTTTGTGTTTCCAAACTCTCACGAATCCTTATCGA	1140
Db	12782	ATCACTTACGCCAAATTTTCCCTAGTGTTTGTGTTTCCAAACTCTCACGAATCCTTATCGA	12841
QY	1141	ACTTTTATATATACATTTGTAGCATTTGGAAAGTATCTTTGTATGCTTTGCTTAAACT	1200
Db	12842	ACTTTTATATATACATTTGTAGCATTTGGAAAGTATCTTTGTATGCTTTGCTTAAACT	12901
QY	1201	TAGACATCCTTGTCTCTGTTTGTGTTTGTGAACTTGCTTGACTTAAATGAAGTTAAAAAT	1260
Db	12902	TAGACATCCTTGTCTCTGTTTGTGTTTGTGAACTTGCTTGACTTAAATGAAGTTAAAAAT	12961
QY	1261	TTGTAGTTAAAAATAGAAAAATTTTACTAAATTTGAGTTTCGATTAATCATAGTCTAGATAAT	1320
Db	12962	TTGTAGTTAAAAATAGAAAAATTTTACTAAATTTGAGTTTCGATTAATCATAGTCTAGATAAT	13021
QY	1321	TTGAAAAAAATTAATAAAATTTTGGAAAAATACATATATGTTTATTTTTTAAAAATAATTTA	1380
Db	13022	TTGAAAAAAATTAATAAAATTTTGGAAAAATACATATATGTTTATTTTTTAAAAATAATTTA	13081
QY	1381	CTAAATTGATAAGTATATAGATTAGTTAGTTTTTTTTTTTTTTTTTTTAAATTTGAAAAACCTC	1440
Db	13082	CTAAATTGATAAGTATATAGATTAGTTAGTTTTTTTTTTTTTTTTTTTAAATTTGAAAAACCTC	13141
QY	1441	ACAAATTATTAATTGAAAGATTCAAAATGCAATGCTTAGTTTTTTTAAAGTTTAATCACCCA	1500
Db	13142	ACAAATTATTAATTGAAAGATTCAAAATGCAATGCTTAGTTTTTTTAAAGTTTAATCACCCA	13201
QY	1501	AACGGATAATTGACCCGAACGACTAATTTCAGTTCGTATACGGGTACAGTCAAAATAACCCG	1560
Db	13202	AACGGATAATTGACCCGAACGACTAATTTCAGTTCGTATACGGGTACAGTCAAAATAACCCG	13261
QY	1561	ACCGGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTCACGGCTTTCGCGA	1620
Db	13262	ACCGGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTCACGGCTTTCGCGA	13321
QY	1621	ATCGCATCACGAGCCTTCCTTCTCTCCTACGCTCTTTTCATCAACCCCTCTCTGCGAATC	1680
Db	13322	ATCGCATCACGAGCCTTCCTTCTCTCCTACGCTCTTTTCATCAACCCCTCTCTGCGAATC	13381
QY	1681	CAATTCTGGTATTTACGATTCTCTTTTCGACTACGCCCAATCGCAATTCAGCCCGTACGA	1740
Db	13382	CAATTCTGGTATTTACGATTCTCTTTTCGACTACGCCCAATCGCAATTCAGCCCGTACGA	13441
QY	1741	ATTTTATATATTGATTATGATCTTGATCCCTTACCTTTTCGTGCTTTTCCTTGATCGAATTT	1800
Db	13442	ATTTTATATATTGATTATGATCTTGATCCCTTACCTTTTCGTGCTTTTCCTTGATCGAATTT	13501
QY	1801	ATGTTGATTTATGCTGCGAAATCAAAATTAATGGATACAGAGTAATTTTGATATATGAT	1860
Db	13502	ATGTTGATTTATGCTGCGAAATCAAAATTAATGGATACAGAGTAATTTTGATATATGAT	13561
QY	1861	GTGTAGGAGCTAAAAATTAGCAACAATACGFAAAGTAATTGAAATCGAAATCATAAACGT	1920
Db	13562	GTGTAGGAGCTAAAAATTAGCAACAATACGFAAAGTAATTGAAATCGAAATCATAAACGT	13621
QY	1921	TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTGATAGCTTGTGGTCAATG	1980
Db	13622	TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTGATAGCTTGTGGTCAATG	13681
QY	1981	GATCAGATAAAAGTTTGTAAAGAAATGACAAAGAAAAATGAGAAGAGTGGCAAGTACGAGG	2040
Db	13682	GATCAGATAAAAGTTTGTAAAGAAATGACAAAGAAAAATGAGAAGAGTGGCAAGTACGAGG	13741
QY	2041	TTGGTCGCACAAATAGGTGAAGGAACCTTTTGCTAAAGGTTAAGTTTTCGAGGAACACAGACA	2100

Db	13742	TTGGTCGCACAAATAGGTGAAGGAACCTTTTGCTAAAGTTAAGTTTTCGAGGAACACAGACA	13801
QY	2101	CTGGTGATAATGTAAGCCCATCAAAATTAATGCGTAAGAGTACAATACTTTAAGAACAGAAATGG	2160
Db	13802	CTGGTGATAATGTAAGCCCATCAAAATTAATGCGTAAGAGTACAATACTTTAAGAACAGAAATGG	13861
QY	2161	TTGATCAGGTATGTTCTGGATTGTTTTTTTACATGGAACCTAAGGTTGTTGCGTCAATGGT	2220
Db	13862	TTGATCAGGTATGTTCTGGATTGTTTTTTTACATGGAACCTAAGGTTGTTGCGTCAATGGT	13921
QY	2221	ATGATCTTTGATTTTCGTTTAAAGCTCTTTTACAGATAAAAAAGAGAGATATCTATAATGAA	2280
Db	13922	ATGATCTTTGATTTTCGTTTAAAGCTCTTTTACAGATAAAAAAGAGAGATATCTATAATGAA	13981
QY	2281	GATTGTTCTGCACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTTCATGCA	2340
Db	13982	GATTGTTCTGCACCCGAACATAGTGAGGTTGTATGAGGTATGTTTGTGTTTTCATGCA	14041
QY	2341	TCTGCGAAATTTTATCTCTGAAGTGTTTTTTTCATCATTTGTTCTTCTGTTGTTTTTGTG	2400
Db	14042	TCTGCGAAATTTTATCTCTGAAGTGTTTTTTTCATCATTTGTTCTTCTGTTGTTTTTGTG	14101
QY	2401	ATTTTCCCGATGTAGGTGTTGGCGAGTCTCTTCGAAAAATATATATAGTTTTTGGAGTTTGTG	2460
Db	14102	ATTTTCCCGATGTAGGTGTTGGCGAGTCTCTTCGAAAAATATATATAGTTTTTGGAGTTTGTG	14161
QY	2461	ACAGGAGGAGAGCTCTTTGATAGAAATGTACGGAACCTTCCATACCTTGTAGGCAGCTCCA	2520
Db	14162	ACAGGAGGAGAGCTCTTTGATAGAAATGTACGGAACCTTCCATACCTTGTAGGCAGCTCCA	14221
QY	2521	TTAGTTAAAAACCTCTCTACTTAATTTTTTAAATATATGAAATCTTTCATGCAGGTTCATAA	2580
Db	14222	TTAGTTAAAAACCTCTCTACTTAATTTTTTAAATATATGAAATCTTTCATGCAGGTTCATAA	14281
QY	2581	AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACCTTCAACAGCTTGTAGATGCTGTTGC	2640
Db	14282	AGGAGGCTTGAAGAAAGTGAGTCTCGGAAATACCTTCAACAGCTTGTAGATGCTGTTGC	14341
QY	2641	TCATTGTCATGCAAGGGTGTTTTACCAACCGTGACCTAAAGGTAAGACGTTGTTTTGTTT	2700
Db	14342	TCATTGTCATGCAAGGGTGTTTTACCAACCGTGACCTAAAGGTAAGACGTTGTTTTGTTT	14401
QY	2701	ACCAATATTCCTCAGAAATATCTCAGTTCGCTTGCAATCCAGACTTGATATTTTGTGTCGC	2760
Db	14402	ACCAATATTCCTCAGAAATATCTCAGTTCGCTTGCAATCCAGACTTGATATTTTGTGTCGC	14461
QY	2761	TATGTTATGTTATCTAGCCAGAAATCTTTTACTCGATACAAATCGAAATCTGAAGGTTT	2820
Db	14462	TATGTTATGTTATCTAGCCAGAAATCTTTTACTCGATACAAATCGAAATCTGAAGGTTT	14521
QY	2821	CGGATTTTCGGACTCAGTGCAATTCCTCAGGAAAGTAAGTCTCTTATCTCTGCTTCAGCAG	2880
Db	14522	CGGATTTTCGGACTCAGTGCAATTCCTCAGGAAAGTAAGTCTCTTATCTCTGCTTCAGCAG	14581
QY	2881	TCTGCTTACGTGGTCATTAACCTTGTTATATACTCAATCAGGAGTAGAACTTCTGCGTAC	2940
Db	14582	TCTGCTTACGTGGTCATTAACCTTGTTATATACTCAATCAGGAGTAGAACTTCTGCGTAC	14641
QY	2941	CACATGTGGAACCTCCGAACCTATGTAGCTCCAGAGGTACTTTAGTGACAGGTTTACGATGG	3000
Db	14642	CACATGTGGAACCTCCGAACCTATGTAGCTCCAGAGGTACTTTAGTGACAGGTTTACGATGG	14701
QY	3001	TTCAGCAGCTGATATTTTGGTCTTCGGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT	3060
Db	14702	TTCAGCAGCTGATATTTTGGTCTTCGGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT	14761
QY	3061	ACCTTTTTCGGAGACCGATCTTCCAGGGTTGTACAGAAAGTAAGTAAACATATCTTTTCGG	3120
Db	14762	ACCTTTTTCGGAGACCGATCTTCCAGGGTTGTACAGAAAGTAAGTAAACATATCTTTTCGG	14821
QY	3121	GAAGAAATCATGAATTCCTTGTTCATGGCTTTTGTCAAACCGTTTATTTGATTTGGTTTTGC	3180



Db	14822	GAAGAAATCATGAATTCCTTGTTCATGGCTTTTGTCAAACCGTTTATTATGATTTGGTTTGC	14881
QY	3181	AAATTTCACCCCTTAGTTTTGAGCTTTTACACATTTGTTATTACAGATAAATGCAGCAGAGT	3240
Db	14882	AAATTTCACCCCTTAGTTTTGAGCTTTTACACATTTGTTATTACAGATAAATGCAGCAGAGT	14941
QY	3241	TTTCTTGTCCACCGTGGTTTTCCGCAGAAGTGAAGTTTTTAAATACATAGATACTTGGACC	3300
Db	14942	TTTCTTGTCCACCGTGGTTTTCCGCAGAAGTGAAGTTTTTAAATACATAGATACTTGGACC	15001
QY	3301	CCAATCCCAAAACAGTGAGTATTTTGTCTCTCCTAGCTATCAGGTTTTTGGTGAT	3360
Db	15002	CCAATCCCAAAACAGTGAGTATTTTGTCTCTCCTAGCTATCAGGTTTTTGGTGAT	15061
QY	3361	ATTTAAATGTTCTAGTAATTATATCTGTTTATCTATTAATGTTTCTCAATTAGAGCAGCGT	3420
Db	15062	ATTTAAATGTTCTAGTAATTATATCTGTTTATCTATTAATGTTTCTCAATTAGAGCAGCGT	15121
QY	3421	ATTCAAATTCAGGAATCAAGAAAGATCCTTGGTTCAGATTAATTTATGTCCTATACGA	3480
Db	15122	ATTCAAATTCAGGAATCAAGAAAGATCCTTGGTTCAGATTAATTTATGTCCTATACGA	15181
QY	3481	GCAAGGGAAGAAAGAAAGTGAATTTGGATGATATTTCGTGCAGTTTTTTTGATGGAATTGAG	3540
Db	15182	GCAAGGGAAGAAAGAAAGTGAATTTGGATGATATTTCGTGCAGTTTTTTTGATGGAATTGAG	15241
QY	3541	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCCAATTTAAATGTTAAGAT	3600
Db	15242	GTTTGTGCTTCGCCCTTCATTATATGCTCTTTGCTGGTCAATTCCAATTTAAATGTTAAGAT	15301
QY	3601	CTCTTAGGAACGTTTGGATGACCAAGAAGTGTCTTGGATGATATTTCGTGCAGTTTTTTTGATGGAATTGAG	3660
Db	15302	CTCTTAGGAACGTTTGGATGACCAAGAAGTGTCTTGGATGATATTTCGTGCAGTTTTTTTGATGGAATTGAG	15361
QY	3661	AGGCATGTGTTAGTTACCAAAACCTGTAAACTGTTTCTTTATTCAATTCGCCAAACCATAG	3720
Db	15362	AGGCATGTGTTAGTTACCAAAACCTGTAAACTGTTTCTTTATTCAATTCGCCAAACCATAG	15421
QY	3721	ACCTTAGGAAGACTTAGATTTTACAAGAGAAATTCCTATTTCTCGACCAAAAACCCCTAGACA	3780
Db	15422	ACCTTAGGAAGACTTAGATTTTACAAGAGAAATTCCTATTTCTCGACCAAAAACCCCTAGACA	15481
QY	3781	AAATCCAGAATACCCCTAGGGCTAAATTACAATGTTTCATGTACCTATCAATATATATCTCC	3840
Db	15482	AAATCCAGAATACCCCTAGGGCTAAATTACAATGTTTCATGTACCTATCAATATATATCTCC	15541
QY	3841	TGCTAAGATTGCTTAGCTTTGGTATAGCCCTAGATATATATAGATACCGTAATTTCTAAT	3900
Db	15542	TGCTAAGATTGCTTAGCTTTGGTATAGCCCTAGATATATATAGATACCGTAATTTCTAAT	15601
QY	3901	GCATATTTAGAAATGTTTTACTTAAACTCAGGCTCTTGCTCTTCTTCTAAAACCTGTACTTCAA	3960
Db	15602	GCATATTTAGAAATGTTTTACTTAAACTCAGGCTCTTGCTCTTCTTAAAACCTGTACTTCAA	15661
QY	3961	TTGTTAAACTAAACCTCAGTATCTGTCTTAGCTAAAAGTTACTTTTACTTGTTTTTCATT	4020
Db	15662	TTGTTAAACTAAACCTCAGTATCTGTCTTAGCTAAAAGTTACTTTTACTTGTTTTTCATT	15721
QY	4021	AAGTTGACCTGTCAATTGCACCTTGTTCACAGGCGAGTTATGTAGCGGAGAATGTAGAGAG	4080
Db	15722	AAGTTGACCTGTCAATTGCACCTTGTTCACAGGCGAGTTATGTAGCGGAGAATGTAGAGAG	15781
QY	4081	AAATGATGAAGGGCCCTGATGATGAATGCCTTTGAGATGATTACCTTATCACAAGGCTT	4140
Db	15782	AAATGATGAAGGGCCCTGATGATGAATGCCTTTGAGATGATTACCTTATCACAAGGCTT	15841
QY	4141	AAATTTATCTGCACCTATTTGACAGGCGACAGGTAGTACCTGATTTTCTATTACTGGTCAT	4200
Db	15842	AAATTTATCTGCACCTATTTGACAGGCGACAGGTAGTACCTGATTTTCTATTACTGGTCAT	15901
QY	4201	AGAGATCTCCATTTTCCAATAAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT	4260
Db	15902	AGAGATCTCCATTTTCCAATAAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT	15961

QY	4261	TGACTGCCCTTATGATGCTGTGTTCTTTAGTTTGTATTATAAATAACTATAAGTTCATTAGATGA	4320
Db	15962	TGACTGCCCTTATGATGCTGTGTTCTTTAGTTTGTATTATAAATAACTATAAGTTCATTAGATGA	16021
QY	4321	TTGGTTGCATGGCATTAGTAGATACAAATGGAATCCAAAATGTTTCCCTGCATATTGATGGC	4380
Db	16022	TTGGTTGCATGGCATTAGTAGATACAAATGGAATCCAAAATGTTTCCCTGCATATTGATGGC	16081
QY	4381	TGATCCCTTTGATCTCGCAGGATTTTGTAAAGGCAAAACCCGTTTTTCTCGAAGGGA	4440
Db	16082	TGATCCCTTTGATCTCGCAGGATTTTGTAAAGGCAAAACCCGTTTTTCTCGAAGGGA	16141
QY	4441	ACCTAGTCAGATAAATTGCTAACATTGAGGCTGTAGCGAACTCAATGGGTTTTTAAGTCTCA	4500
Db	16142	ACCTAGTCAGATAAATTGCTAACATTGAGGCTGTAGCGAACTCAATGGGTTTTTAAGTCTCA	16201
QY	4501	TACACGAAACTTCAAGGTAACGAATTC	4527
Db	16202	TACACGAAACTTCAAGGTAACGAATTC	16228

RESULT 4  
AC006258/c

LOCUS	AC006258	110469 bp	DNA	linear	PLN 28-DEC-1998			
DEFINITION	Arabidopsis thaliana BAC F18G18 from chromosome V near 60.5 cM, complete sequence.							
ACCESSION	AC006258							
VERSION	AC006258.1	GI:4063735						
KEYWORDS	HTG.							
SOURCE	Arabidopsis thaliana (thale cress)							
ORGANISM	Arabidopsis thaliana							
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.							
	1 (bases 1 to 110469)							
	Huang,E.N., Dedhia,N., de la Bastide,M., Habermann,K., Matero,A., Preston,R., Shekher,M., Rodriguez,M., Nascimento,L., Spiegel,L.A., Schutz,K., Shah,R., Swaby,I., Vil,M.D., O'Shaughnessy,A., Parnell,L.D. and McCombie,W.R.							
AUTHORS	Arabidopsis thaliana BAC F18G18 from chromosome V near 60.5 cM							
TITLE	Unpublished							
JOURNAL	2 (bases 1 to 110469)							
REFERENCE	Huang,E.N., Dedhia,N., de la Bastide,M., Habermann,K., Matero,A., Preston,R., Shekher,M., Rodriguez,M., Nascimento,L., Spiegel,L.A., Schutz,K., Shah,R., Swaby,I., Vil,M.D., O'Shaughnessy,A., Parnell,L.D. and McCombie,W.R.							
AUTHORS	Direct Submission							
TITLE	Submitted (28-DEC-1998) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA							
JOURNAL	BAC F18G18 is assigned to YAC C1C2E3 and maps to near 60.5 cM on the Lister & Dean RI map. Position 1 of F18G18 is oriented toward the telomere and position 110469 is oriented toward the centromere. For more information on the mapping, sequencing and annotation of F18G18, please see							
COMMENT	http://www.cshl.org/arabweb/F18G18-titlepage.html. A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan (http://CCR-081.mit.edu/GENSCAN.html), MZEF (http://www.cshl.org/genefinder) and GRAIL (http://compbio.ornl.gov/tools/index.shtml) and with splice sites predicted by NetPlantGene (http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html). Alternate exons not used in building the gene models are presented on the web pages associated with F18G18. Genes are numbered according to the scheme BAC.gene_number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (http://expasy.hcuge.ch/sprot/prosite.html) and Pfam (http://pfam.wustl.edu/) libraries. A description of these categories can be found at							



http://muntjac.mips.biochem.mpg.de/arabi/. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F18G18, please direct email to Larry Parnell at parnell@cshl.org.

FEATURES	Location/Qualifiers
	1..110469
source	/organism="Arabidopsis thaliana"
	/mol_type="genomic DNA"
misc_feature	/cultivar="Columbia"
	/db_xref="taxon:3702"
repeat_region	/chromosome="V"
	/map="near 60.5 cm"
repeat_region	1..16462
	/note="overlap with F21J6 from position 94219 to 110680"
repeat_region	7777..9874
	/note="function=chloroplast_repeat; similar to region between trna-Pro and ribosomal L12 genes, GenBank accession number X68046"
repeat_region	/rpt_type=dispersed
	11865..12483
repeat_region	/note="function=MITES-like; 22 copies of a 28-bp repeat; similar to but distinct from a 28-bp MITEs-like element found in T32N15, GenBank accession number AC002534"
	/rpt_type=tandem
repeat_region	/rpt_unit=TTTCCCGCCAAAACGTAAYCCGTRA
	13331..13358
repeat_region	/note="function=putative_microsatellite"
	/rpt_type=tandem
repeat_region	/rpt_unit=AT
	complement(27355..27387)
repeat_region	/note="function=putative_microsatellite"
	/rpt_type=tandem
repeat_region	/rpt_unit=ATG
	57299..60193
repeat_region	/gene="F18G18.15"
	/note="encodes cyclin 3a, GenBank accession number Z31589"
repeat_region	/evidence=experimental
	join(57299..57592,57676..57817,57883..58022,58129..58221,58337..58470,58663..58738,58836..58912,59009..59042,59129..59278,59361..59495,59576..59635)
repeat_region	/gene="F18G18.15"
	/note="functional catalog ID=03.22; functional catalog ID=10.99"
repeat_region	/codon_start=1
	/evidence=experimental
repeat_region	/product="cyclin 3a"
	/protein_id="AAC98445.1"
repeat_region	/db_xref="GI:4063736"
	/translation="MHRASSKHTNAKKEAISTKIRDNNVVRTRSRAKALGVSNPSK PAFKETKRVARPSNKRMSDNTVCNQKRRRAVLKDVNTIAESIISTEGNVKVAACKR GGKETQIEEDGLVDVDGEKSLAEDLSKIRMVESLDASASKQLVDCAEEDRSVDTD CVQIVDIDSGVQDPQFCSLYASIDYSINVAELESPPSYMVQVQVDIDPTMRGILI DMLVEVSEYKLVSDTLVLTNLDIFMSHNYIEKQLQLLGITCMLIAASKYEEISAP RLEEFCTIDNTYTRLEVLSMEIKVLNSLHFRLSVPTTKTFLRRFIRAAQASDKVPLI EMEYLANVFAELTLTEYTFRLFPLSLAASAVFLARWTLDOQSNHPNQTLOHYTRYET SALKNTVLAMEELQNTSGSTLIAHTKYNQKFKRVATLTSPEVNTLFSR"
repeat_region	89474..89507
	/note="function=putative_microsatellite"
repeat_region	/rpt_type=tandem
	/rpt_unit=AT
repeat_region	37147 a 18606 c 18142 g 36574 t
	BASE COUNT 37147 a 18606 c 18142 g 36574 t
repeat_region	ORIGIN
	Query Match 8.9%; Score 459.4; DB 8; Length 110469;
repeat_region	Best Local Similarity 84.7%; Pred. No. 1.1e-62;
	Matches 541; Conservative 0; Mismatches 91; Indels 7; Gaps 2;
repeat_region	2 TTTTGGCGGAAATCTCGGGTTTACGTTTTCGGGGGAAATCTCGTGTTCACGTTTTT
	QY 61

Db	12018	TTTTGGCGTAAAGTTACGGGTTTTTCGTTTTTGGCGGAAATCTCGGTTTACGTTTTT	11959
	62	GGCGGGAAATCTCGGGTTTACGGTTTTTTCGCGAGAAATCACGGGTTACTTTTTTTC	121
Db	11958	GACGGGAAATCTCGGGTTTTTGTGTTTTTGGAGAAATCATGGATTAC-GTTTTAGC	11900
	122	GGGAAATCACGGATTATATGTTTTTTCGTTGGGAAATACGAGTTACTTTTTCTCAAT	181
Db	11899	GGGAAATCACGGATTATGTTTTTTCGTTGGGAAATACGAGTTACTTTTTCTCAAT	11840
	182	TCATTGCTTGTATATTTAAGAAATTTGGAAAAATATTAATTTTATAAATGGTTAGAT	241
Db	11839	TCATCGCTTGTATATTTAAGAAATTTGGAAAAATATTAATTTTATAAATGGTTAGAT	11780
	242	GTGTTGGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTTAGTTGGTTTTTCAATTT	301
Db	11779	GTGTTGGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTTAGTTGGTTTTTCAATTT	11720
	302	ACAAATTTGATGGGTTAATTTGGATAAACCATGGAAACCATTAACCATTAACCACTA	361
Db	11719	ACAAATTTGATGGGTTAATTTGGATAAACCATGGAAACCATTAACCATTAACCACTA	11660
	362	CATTTTACTCATCAAACTCAATTTGACTCATCAACTCATTTGACTCATCACTTTGAGT	421
Db	11659	CATTTTACTCATCAAACTCAATTTGACTCATCAACTCATTTGACTCATCACTTTGAGT	11600
QY	422	CAAAATTTCAACTCATTTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT	481
	11599	CAAAATTTCAACTCATTTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT	11540
QY	482	TGACCCATTTTGACACCCCTACATATGATCATATAAGTTAATAATAAAAAATTAATTA	541
	11539	TGACCCATTTTGACACCCCTA-----ACAATATGTAAATATGTAAACCATGTTCT	11486
QY	542	AACTTTTTAACGGAATTTGTTTGTAAAGTTTCATTACGTTATTTATATAAAAAACATA	601
	11485	CCAGTGAGAAATCAATCGTCTTTTAAATATGTATGAATCATGGAGCTTCAATAATTAG	11426
QY	602	TGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATA 640	
	11425	TAAGTACATAATTACCTACCAATTTGAATTTTAAATA 11387	

RESULT 5

AC006259/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

BAC F21J6 is assigned to YAC CIC2E3 and maps to near 60.5 cm on the Lister & Dean RI map. Position 1 of F21J6, the T7 end, is oriented toward the telomere and position 110680, the SP6 end, is oriented toward the centromere. For more information on the mapping, sequencing and annotation of F21J6, please see <http://www.cshl.org/arabweb/F21J6-titlepage.html>. A graphic view of our annotation is also available at this url. Gene models are built with exons predicted by GenScan (<http://CCR-081.mit.edu/GENSCAN.html>), MZEF (<http://www.cshl.org/genefinder>) and GRAIL (<http://compbio.ornl.gov/tools/index.shtml>) and with splice sites predicted by NetPlantGene (<http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html>). Alternate exons not used in building the gene models are presented on the web pages associated with F21J6. Genes are numbered according to the scheme BAC.gene\_number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Protein sequences ecoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite (<http://expasy.hcuge.ch/sprot/prosite.html>) and Pfam (<http://pfam.wustl.edu/>) libraries. A description of these categories can be found at <http://muntjac.mips.biochem.mpg.de/arabi/>. Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence concerning the annotation of F21J6, please direct email to Larry Parnell at [parnell@cshl.org](mailto:parnell@cshl.org).

FEATURES

source

1. 110680  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="V"  
/map="near 60.5 cm"

misc\_feature

1. .24515  
/note="overlap with T11H3, GenBank accession number AC005964, from position 67404 to 91916"  
<1. .717  
/gene="F21J6.1"

gene

/note="encodes putative P450; entire gene encoded on T11H3, GenBank accession number AC005964; genomic copy of EST Z33677; genomic copy of EST Z33963; gene model last edited on 16 Dec 98"  
/evidence=not\_experimental  
join(<1. .11,103. .717)  
/gene="F21J6.1"

CDS

/note="functional catalog ID=98"  
/codon\_start=3  
/evidence=not\_experimental  
/product="putative P450"  
/protein\_id="AAC98443.1"  
/db\_xref="GI:4063733"

/translation="ILFNILNAGIDTSAQVMTWMTYLISNPRVMKKAQAEVREVIKN  
KDDIIEEDIERLEYLKMVVKETFRVLPVLLIPREASKDVKIGGYDIPKKTWIVHNI  
WAIHRNPVWKDPEAFIPERFMDNQIDYKGLNFELPFGSGRMCPCGIGMGMAVHLT  
LINLLYRFDWKLPEGMEVEDVLEESYGLVCPKKVPLQLIPVLTQWS"  
complement(1310. .1353)  
/note="function=putative\_microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=AT  
4808. .6391

repeat\_region

gene

/gene="F21J6.2"  
/note="encodes putative P450; gene model last edited on 16 Dec 98"  
/evidence=not\_experimental  
join(4808. .5683,5777. .6391)  
/gene="F21J6.2"  
/note="functional catalog ID=98"  
/codon\_start=1

CDS

/evidence=not\_experimental  
join(4808. .5683,5777. .6391)  
/gene="F21J6.2"  
/note="functional catalog ID=98"  
/codon\_start=1

/evidence=not\_experimental  
/product="putative P450"  
/protein\_id="AAC98444.1"  
/db\_xref="GI:4063734"  
/translation="MSLWYIIIVAFVFFSSMIIVRIIRKTKKNLPPGPPRLPIIGNLHQ  
LGSXPHRSMFKLSETYGPLMSLKFGSVSTVASTPETVKEVLTDFDVECCSRPNMTYP  
ARVTYNLKDLCFSPYSKYWRVRKMTVVELYTAKRVQSFQHTRKEEVAALVDFIKQAA  
SLEKPVNLKKMLKSGSVICRVAFGINLQGSKLENTYEEVIQGTVELVGSFAAADYF  
PVVGRIIDRITGLHSKCEKLFKAMDAFFDQSIKHHLEDEIHKDDIIDLLKKMERGETT  
LGEFQLTRDHTKGILANILNAGIDTSAQVMTWMTYLISNPRVLKKAQAEVREVIKHK  
DDIIEEDIERLQYLKMVVKETFRINPLVPLLIIPREASKDVKIGGYNI PKKTWIVHNIW  
AIHRNPVWKDPEAFIPERFMDSQIDYKGLNFELLPFGSGRRICPGIGMGMAVHLTL  
INLLYRFDWKLPEGMKVADVLEESYGLVCPKKIPLQLIPVLTQWT"  
23757. .24702  
/gene="F21J6.6"  
/note="encodes ZFP3 zinc finger protein; genomic copy of cDNA L39646"  
/evidence=experimental  
23757. .24702  
/gene="F21J6.6"  
/note="mRNA sequence based on a comparison to cDNA L39646"  
23757. .23810  
/gene="F21J6.6"  
/note="5'-UTR sequence based on a comparison to cDNA L39646"  
23811. .24518  
/gene="F21J6.6"  
/note="functional catalog ID=98"  
/codon\_start=1  
/evidence=experimental  
/product="ZFP3 zinc finger protein"  
/protein\_id="AAC98442.1"  
/db\_xref="GI:4063732"  
/translation="MDASIVSSSTAFPYQDSLNSQSIEDERDVHNSSHELNLIDCID  
TTSIVNESTTSTEQKLFSCNYCQRTFYSSQALGCHQNAHKRERTLAKRGORMAASASA  
FGHPYGFSPLPFPGQYNNHRSLSLGIQAHSISHKLSYNGFGHYGQINWSRLPFDQOPA  
IGKFPMDNFHHHHQMMMAPSVNSRNIDSPSNTGRVLEGSPTLEQWHGDKGLLL  
STSHHEEQKLDLSLKL"  
24519. .24702  
/gene="F21J6.6"  
/note="3'-UTR sequence based on a comparison to cDNA L39646"  
38122. .38161  
/note="function=putative\_microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=AT  
48069. .48103  
/note="function=putative\_microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=A  
66245. .66277  
/note="function=putative\_microsatellite"  
/rpt\_type=tandem  
/rpt\_unit=AAC  
72338. .74677  
/gene="F21J6.18"  
/note="encodes KNOTTED-like homeobox protein KNAT3;  
genomic copy of cDNA X92392; genomic copy of EST T44158;  
genomic copy of EST T44387"  
/evidence=experimental  
join(72338. .73175,73451. .73571,73660. .73844,73922. .74053,  
74132. .74268,74363. .74677)  
/gene="F21J6.18"  
/note="mRNA sequence based on a comparison to cDNA X92392"  
72338. .72494  
/gene="F21J6.18"  
/note="5'-UTR sequence based on a comparison to cDNA X92392"  
join(72495. .73175,73451. .73571,73660. .73844,73922. .74053,  
74132. .74268,74363. .74402)  
/gene="F21J6.18"  
/note="functional catalog ID=40.02; functional catalog ID=50.02"

		/codon_start=1 /evidence=experimental /product="KNAT3 homeodomain protein" /protein_id="AAC98441.1" /db_xref="GI:4063731" /translation="MAFHHNLSQDLSFNHFTDOHPPPPPOPPPPPOQQQHFQEAAPP PNWLNTALLRSSDNNNNFLNLTATANTTTASSSDSPSSAAAAAANQWLSSSSFLQ RNNNNASIVGDIDVTGGADTMIQGMKGTGGKNDGGGATAADGVVSWONARHK AEILSHPLYEQLLSAHVACLRATPVDQLPRIDAQLAQSHVYAKYSALGAAAQGLVG DDKELDPWTHYVLLLCSEKQEQLOQHVRVHAMEAVMACWEIEQSLQSLTGVSPGEGMG ATMSDDEQVESDANMFDGGLDVLFGLPIPTESERSLMERVQELKHELKQGYKEK IVDIRREILRKRRAGKLPGDTTSVLKAWQSHSKWPYPTEEDKARLVQETGLQLKQIN NWFINQRKNWHSNPSSSTVLKNRKRKSNAGDNSGRERFA" complement(72743..72760) /note="function=trinucleotide_repeat; amino_acid=Ala" /rpt_type=tandem /rpt_unit=GCG 74403..74677 /gene="F21J6.18" /note="3'-UTR sequence based on a comparison to cDNA X92392" complement(80910..82988) /note="function=chloroplast_repeat; similar to region between tRNA-Pro and ribosomal protein L12 genes, GenBank accession number X68046" /rpt_type=dispersed 94219..110680	
repeat_region			
3'UTR			
repeat_region			
misc_feature			
Query Match Best Local Similarity 84.7%; Score 459.4; DB 8; Length 110680; Matches 541; Conservative 0; Mismatches 91; Indels 7; Gaps 2;			
QY	2	TTTTGGCGGAAAAATCTCGGGTTTACGTTTTGGCGGAAAAATCTCGGTTTACGTTTTT	61
Db	106236	TTTTGGCGGTAAGTTACGGGTTTTCGTTTTGGCGGAAAAATCTCGGTTTACGTTTTT	106177
QY	62	GGCGGAAAAATCTCGGGTTTACGTTTTTGGAGAAAAATCACGGGTTTACTTTTTTTC	121
Db	106176	GACGGGAAAAATCTCGGGTTTTCGTTTTTGGAGAAAAATCATGGATTAC-GTTTTTAC	106118
QY	122	GGGAAAAATCACGGATTATGTTTTTTTGTGGAAAAAATTACGAGTTTACTTTTCTCAAT	181
Db	106117	GGGAAAAATCACGGATTATGTTTTTTTGTGGTGA AAAAATTACGAGTTTACTTTTCTCAAT	106058
QY	182	TCATTGCTTGATATTTAAGAAATTTGGAAAAATATTAAATTTTATTAAATCGTTTAGAT	241
Db	106057	TCATCGCTTGATATTTAAGAAATTTGGAAAAATATTAAATTTTATTAAATCGTTTAGAT	105998
QY	242	GTGTTGGTTAAACCTAAATTGGCATTGGTTTAGAGATTTTAGTTGGTTTATTCAATTTT	301
Db	105997	GTGTTGGTTAAACTTAAATTGGCATTGGTTTAGAGATTTTAGTTGGTTTATTCAATTTT	105938
QY	302	ACAAAATTTGATGGGTTAATTGGATAAACCATGGAAACCATTAACCAATACAACCTAACT	361
Db	105937	ACAAAATTTGATGGGTTAATTGGGTAAACCATTGA AACCATTAACCAATACAACCCAACT	105878
QY	362	CATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACTCATCAACTCATTTGAGT	421
Db	105877	CATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACCCATCAACTCATTTGAGT	105818
QY	422	CAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACCCCATGAATTT	481
Db	105817	CAAAAATTTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACTCATGAATTT	105758
QY	482	TGACCCATTTTGACACCCCTACATATGATCATATAAGTTAATAATCAAAAATTACTATTGAT	541
Db	105757	TGACCCATTTTGACACCCCTA-----ACAATATGTAATAATGTAAAACATGTTCTCTAGT	105704
QY	542	AACTTTTTAAACGGAATTGTTTTGTAAAGTTTTCATTTTACGTTATTATATAAAAAACATAA	601
Db	105703	CCAGTGAGAAATTCATCGTCTTTTAAAAATGTATGAATCATGGAGCTTCAATAATTAGTGG	105644
QY	602	TGCAAAAGTACTAATGTATAGTTATTTTATTTTTAAATA	640

Db	105643	TAAGTACATAATTACCTACCAATATTGAATTTTAAATA	105605
RESULT 6	AC007399/c		
LOCUS	AC007399	105223 bp	DNA linear PLN 26-APR-1999
DEFINITION	Arabidopsis thaliana BAC F14I23 from chromosome V near 69 cM, complete sequence.		
ACCESSION	AC007399		
VERSION	AC007399.1	GI:4680765	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1	(bases 1 to 105223)	
AUTHORS	Habermann,K., Nascimento,L., Vil,M.D., Matero,A., Rodriguez,M., Shah,R., Swaby,I., Shekher,M., O'Shaughnessy,A., Huang,E.N., Spiegel,L.A., Schutz,K., Parnell,L.D., Preston,R.R., See,L.H., Dedhia,N.N. and McCombie,W.R.		
TITLE	Arabidopsis thaliana BAC F14I23 from chromosome V near 69 cM		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 105223)	
AUTHORS	Habermann,K., Nascimento,L., Vil,M.D., Matero,A., Rodriguez,M., Shah,R., Swaby,I., Shekher,M., O'Shaughnessy,A., Huang,E.N., Spiegel,L.A., Schutz,K., Parnell,L.D., Preston,R.R., See,L.H., Dedhia,N.N. and McCombie,W.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (26-APR-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724		
COMMENT	BAC F14I23 maps to near 69 cM on the Lister & Dean RI map and is assigned to YAC CIC12F8. Position 1 of F14I23 is oriented toward the north telomere centromere and position 105223 is oriented toward the centromere. For more information on the mapping sequencing and annotation of F14I23, please see <a href="http://www.cshl.org/arabweb/F14I23-titlepage.html">http://www.cshl.org/arabweb/F14I23-titlepage.html</a> . A graphic view of our annotation will also be available at this url. Gene models are built with exons predicted by Genscan ( <a href="http://CCR-081.mit.edu/GENSCAN.html">http://CCR-081.mit.edu/GENSCAN.html</a> ), MZEF ( <a href="http://www.cshl.org/genefinder">http://www.cshl.org/genefinder</a> ) and GRAIL ( <a href="http://compbio.ornl.gov/tools/index.shtml">http://compbio.ornl.gov/tools/index.shtml</a> ) and with splice sites predicted by NetPlantGene ( <a href="http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html">http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html</a> ). Genes are numbered according to the scheme BAC.gene_number. Typically, these numbers progress from 1 upwards as one moves from position 1 of the BAC. Hypothetical proteins are those having no EST matches and similarity only to other hypothetical proteins; predicted proteins have EST matches but cannot be assigned a function based on current composition and nomenclature within the nr protein database. Protein sequences encoded by the genes are assigned to a functional category with the aid of similarity searches and comparison to the Prosite ( <a href="http://expasy.hcuge.ch/sprot/prosite.html">http://expasy.hcuge.ch/sprot/prosite.html</a> ) and Pfam ( <a href="http://pfam.wustl.edu/">http://pfam.wustl.edu/</a> ) libraries. A description of these categories can be found at <a href="http://muntjac.mips.biochem.mpg.de/arabi/">http://muntjac.mips.biochem.mpg.de/arabi/</a> . Genomic repeats are typically located by TBLASTX analysis and an attempt is made to classify the function of each repeat as either transposon, putative microsatellite, LINE, direct repeat, centromeric repeat, etc.		
FEATURES	If you have any questions or confirmatory or contradictory evidence concerning the annotation of F14I23, please direct email to Larry Parnell at <a href="mailto:parnell@cshl.org">parnell@cshl.org</a> .		
source	Location/Qualifiers		
	1. .105223		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/cultivar="Columbia"		
	/db_xref="taxon:3702"		
	/chromosome="V"		
	/map="near 69 cM"		









IQVQSFFTALKALSRKEVHYGYSATTEQATCVPSARRRCTPQFFFAANTCFAKIAFQS  
GEVFLHIAIIFVVEYLDVERM"  
1824. .2316  
/gene="T7H20\_10"  
/number=1  
2317. .2402  
/gene="T7H20\_10"  
/number=1  
2403. .2479  
/gene="T7H20\_10"  
/number=2  
2480. .2559  
/gene="T7H20\_10"  
/number=2  
2560. .2721  
/gene="T7H20\_10"  
/number=3  
2722. .2834  
/gene="T7H20\_10"  
/number=3  
2835. .2918  
/gene="T7H20\_10"  
/number=4  
2919. .3086  
/gene="T7H20\_10"  
/number=4  
3087. .3251  
/gene="T7H20\_10"  
/number=5  
3252. .3301  
/gene="T7H20\_10"  
/number=5  
3302. .3406  
/gene="T7H20\_10"  
/number=6  
3407. .3512  
/gene="T7H20\_10"  
/number=6  
3513. .3668  
/gene="T7H20\_10"  
/number=7  
4027. .5664  
/gene="T7H20\_20"  
complement(join(4027. .4332,4414. .4581,4673. .4876,  
4955. .5080,5179. .5283,5518. .5664))  
/gene="T7H20\_20"  
complement(join(4027. .4332,4414. .4581,4673. .4876,  
4955. .5080,5179. .5283,5518. .5664))  
/gene="T7H20\_20"  
/note="similarity to putative protein At2g30530 -  
Arabidopsis thaliana, EMBL:U93215"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB82970.1"  
/db\_xref="GI:7340671"  
/translation="MAYRRRQGIKGFATFKEEVDRLPPESTITAVKDRSPPARSSPADP  
QPRKNFTTEPKGLWGVIQAQKASVIEDDKSSDRSTTASQSRFSYLSDEGFKKMDNPK  
LRRGLDLTSSLNQIGDTFEKAFEDGRTLVENKTADIIQTRKQLQTRRRGTGGEDENQ  
NQSYGVSSSWKKSPQPMQLNHIHETQLKASRDVAMATAAKAKLLILRELKTVKADLA  
FAKERAQLEENKHLREHREKSGNPADEDLIRLQLESLLAEKARLAHENS VYAREN  
RFLRIVEYHQLTMQDVVYIDEGSEEVTVQSPFVSTILMTSSPADRSQSPPSQEMIKEM  
LVVAESSVCASVSVS"  
complement(4027. .4332)  
/gene="T7H20\_20"  
/number=1  
complement(4333. .4413)  
/gene="T7H20\_20"  
/number=1  
complement(4414. .4581)  
/gene="T7H20\_20"  
/number=2  
complement(4582. .4672)  
/gene="T7H20\_20"

IQVQSFFTALKALSRKEVHYGYSATTEQATCVPSARRRCTPQFFFAANTCFAKIAFQS  
GEVFLHIAIIFVVEYLDVERM"  
1824. .2316  
/gene="T7H20\_10"  
/number=1  
2317. .2402  
/gene="T7H20\_10"  
/number=1  
2403. .2479  
/gene="T7H20\_10"  
/number=2  
2480. .2559  
/gene="T7H20\_10"  
/number=2  
2560. .2721  
/gene="T7H20\_10"  
/number=3  
2722. .2834  
/gene="T7H20\_10"  
/number=3  
2835. .2918  
/gene="T7H20\_10"  
/number=4  
2919. .3086  
/gene="T7H20\_10"  
/number=4  
3087. .3251  
/gene="T7H20\_10"  
/number=5  
3252. .3301  
/gene="T7H20\_10"  
/number=5  
3302. .3406  
/gene="T7H20\_10"  
/number=6  
3407. .3512  
/gene="T7H20\_10"  
/number=6  
3513. .3668  
/gene="T7H20\_10"  
/number=7  
4027. .5664  
/gene="T7H20\_20"  
complement(join(4027. .4332,4414. .4581,4673. .4876,  
4955. .5080,5179. .5283,5518. .5664))  
/gene="T7H20\_20"  
complement(join(4027. .4332,4414. .4581,4673. .4876,  
4955. .5080,5179. .5283,5518. .5664))  
/gene="T7H20\_20"  
/note="similarity to putative protein At2g30530 -  
Arabidopsis thaliana, EMBL:U93215"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB82970.1"  
/db\_xref="GI:7340671"  
/translation="MAYRRRQGIKGFATFKEEVDRLPPESTITAVKDRSPPARSSPADP  
QPRKNFTTEPKGLWGVIQAQKASVIEDDKSSDRSTTASQSRFSYLSDEGFKKMDNPK  
LRRGLDLTSSLNQIGDTFEKAFEDGRTLVENKTADIIQTRKQLQTRRRGTGGEDENQ  
NQSYGVSSSWKKSPQPMQLNHIHETQLKASRDVAMATAAKAKLLILRELKTVKADLA  
FAKERAQLEENKHLREHREKSGNPADEDLIRLQLESLLAEKARLAHENS VYAREN  
RFLRIVEYHQLTMQDVVYIDEGSEEVTVQSPFVSTILMTSSPADRSQSPPSQEMIKEM  
LVVAESSVCASVSVS"  
complement(4027. .4332)  
/gene="T7H20\_20"  
/number=1  
complement(4333. .4413)  
/gene="T7H20\_20"  
/number=1  
complement(4414. .4581)  
/gene="T7H20\_20"  
/number=2  
complement(4582. .4672)  
/gene="T7H20\_20"

/number=2  
complement(4673. .4876)  
/gene="T7H20\_20"  
/number=3  
complement(4877. .4954)  
/gene="T7H20\_20"  
/number=3  
complement(4955. .5080)  
/gene="T7H20\_20"  
/number=4  
complement(5081. .5178)  
/gene="T7H20\_20"  
/number=4  
complement(5179. .5283)  
/gene="T7H20\_20"  
/number=5  
complement(5284. .5517)  
/gene="T7H20\_20"  
/number=5  
complement(5518. .5664)  
/gene="T7H20\_20"  
/number=6  
5555. .8036  
/gene="T7H20\_30"  
5555. .8036  
/gene="T7H20\_30"  
/note="similarity to COP1-interacting protein CIP8 -  
Arabidopsis thaliana, EMBL:AF162150"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB82971.1"  
/db\_xref="GI:7340672"  
/translation="MESQGPSNGHPMDISSNIPSSACNLGCRVVQASDDLEIPSLC  
VDCKFLLEDFTGPAPPLARRQTI RRRMRRRRTTRHDSSES VGVGDLSSQQFTHLI  
SIARQSLSTVQASEDLRDDLR TSSHTTPSGSTRWRLFSESDDDFDFNGETESNASFS  
LYRFPHADNDTISFSAYGESDASTDRHADIFVQDDRSDDIDFTDIDPMRAGLNQWN  
SDEEDREWEEGAGPSGVAGTRYRNYLASPSESYSSMTFDSPELERSFRQRIIERHRS  
LSRNIFTGLEDLDFSPYAANVG DYLDERGFDELLEQAESDNSRRGAPPASVSCVRNL  
PRVIAAEHV MKGLVCAICKELFSLRNETTQLPCLHLVHAHCIVPWL SARNSCPLCRY  
ELPTDDKDYEGKRNVLVDS SSSDDGTESGEEYVERGESESGVSRVSRGRWLFL  
AAAPVVLVGVV LAMWLS SPHRRGIAISHSQRENRTRWLPFF"  
5555. .8036  
/gene="T7H20\_30"  
/number=1  
8386. .10613  
/gene="T7H20\_40"  
complement(join(8386. .8696,8788. .8840,8920. .9020,  
9134. .9208,9311. .9541,9843. .9964,10042. .10166,  
10245. .10324,10416. .10613))  
/gene="T7H20\_40"  
complement(join(8386. .8696,8788. .8840,8920. .9020,  
9134. .9208,9311. .9541,9843. .9964,10042. .10166,  
10245. .10324,10416. .10613))  
/gene="T7H20\_40"  
/note="similarity to hypothetical proteins - Arabidopsis  
thaliana"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB82972.1"  
/db\_xref="GI:7340673"  
/translation="MIARILAA LADSMEMPV AAGGCVLGTIKI AVMPIAKVFTMCFL  
GLIMASKYVNILPPSGRKLNLGLVFSLLPCLIFSQLQAVTLQKMLQWWFI PVNVVL  
GTISGSIIGFIVASIVRPYPYFKFTIIQIGVGNIGNVPLVLLAALCRDTSNPFQDSE  
KCSIDGTAYISFGQWVGAILYTYVYQMFAPPPEGFDAEENLAKTL PVDAAPEQVP  
LLTQNFPKDFSPTQDLLPVQSTEPGRGRGVSRKGIAQIFVFLYKCLKQIQVQPAIVA  
SILAMILGAIPFTKKLI FTNGAPLFFFTDSCMLGDAMIPCILLALGGNLINGPSSK  
LGFKTTAAIIIGRLVLPVPVGLGIVTVADKLGFLPADDKMFRFVLLLQHTMPTSVLSG  
AVANLRCCGRESAAVLFVWHIFAIFSMAGWMLVINILF"  
complement(8386. .8696)  
/gene="T7H20\_40"  
/number=1  
complement(8697. .8787)  
/gene="T7H20\_40"

exon		/number=1 complement (8788.. .8840)	
Query Match 8.8%; Score 455.2; DB 8; Length 87581; Best Local Similarity 79.8%; Pred. No. 5.2e-62; Matches 549; Conservative 0; Mismatches 138; Indels 1; Gaps 1;			
QY	1	TTTTTGGCGGAAAAATCTCGGGTTTACGTTTTTGGCGGAAAAATCTCGTGTTCGTTTTT	60
Db	25147	TTTTTGGCGGAAAAATCACGGGTTTACGTTTTTGGTGGAAAAATTACGGGTTTACGTTTTT	25088
QY	61	TGGCGGAAAAATCTCGGGTTTACGTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTG	120
Db	25087	TGGCGGAAAAATTAAAGGTTTACGTTTTTGGCGGAAAAATCACGGGTTTAC-GTTTTTG	25029
QY	121	CGGAAAAATCACGGATTATGTTTTTGGTGGAAAAATTACGAGTTTACTTTTTCTCAAT	180
Db	25028	CGAGAAAAATCACGGGTTTATGTTTTTGGTGGAAAAATTACGAATTTATTTTCTCAAT	24969
QY	181	TTCATTTGCTTGATATTTAAGAAATTTGGAAAAATATTAATTTTATTAAATTTGTTAGA	240
Db	24968	TTCATCGCTTGATATTTAAGAAATTTGGAAAAATATTAATTTTATTAAATTTGTTAGA	24909
QY	241	TGTGTTGGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTTAGTTTATTCAATTT	300
Db	24908	TGTGTTGGTTAAACTTAAATTTGGCATTTGGTTTAGAGATTTTAGTTTATTCAATTT	24849
QY	301	TACAAAAATTTGATGGGTTAATTTGGATAAACCATGGAAACCATTAACCAATFACAAACCTAAC	360
Db	24848	TATAAAATTTGATGGGTTAATTTGGGTAAACCATTTGAAACCATTAACCAATFACAAACCAAC	24789
QY	361	TCATTTTACTCATCAAAACCAATTTGACTCATCACTCACTCATTTGACTCATCACTCATTTGAG	420
Db	24788	TCATTTTACTCATCAAAACCAATTTGACTCATTAACCCATTTGACTCATCACTCATTTGAG	24729
QY	421	TCAAAAATTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGACCCATGAATT	480
Db	24728	TCAAAAATTTTCAACTCATTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGACCCATGAATT	24669
QY	481	TTGACCCATTTTGACACCCCTACATATGATCATATAAGTTAATAATCAAAAATTACTATTGA	540
Db	24668	TTGACCTATTTTGACATCCCTAACTCAAAGCATGGATGGTATGGTATGACCCGTAAGATT	24609
QY	541	TAACTTTTAAACGAATTTGTTTGTGAAGTTTCATTTACGTTATTTATATAAAAAACATA	600
Db	24608	ATACATTGAGAACAAATCAAATTTCTGCAGTTTCATATTATATAATTAGAGAAATCACAAAGA	24549
QY	601	ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGCAAAAATATTACTGTAA	660
Db	24548	TTTAAAAAGATGCGCTAGATTAAATTATGTTAACAAAATTTATAGTCCATAAATTAATCAAAA	24489
QY	661	TACTTTTTCATCTTATCAATATTTTTT 688	
Db	24488	TTGTTGTGAATTTGTGATCAACCATATTT 24461	
RESULT 10			
AC007887/c			
LOCUS		AC007887 158096 bp DNA linear PLN 28-JUN-2000	
DEFINITION		Genomic sequence for Arabidopsis thaliana BAC F1504 from chromosome I, complete sequence.	
ACCESSION		AC007887	
VERSION	AC007887.9	GI:8778333	
KEYWORDS		HTG.	
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM		Arabidopsis thaliana	
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE	1 (bases 1 to 158096)		
AUTHORS	Ecker, J.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-JUN-1999) Arabidopsis thaliana Genome Center,		

REFERENCE AUTHORS TITLE JOURNAL	Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
	2 (bases 1 to 158096)		
	Ecker, J.R.		
REFERENCE AUTHORS	Direct Submission		
	Submitted (04-OCT-1999) Arabidopsis thaliana Genome Center,		
	Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA		
REFERENCE AUTHORS	3 (bases 1 to 158096)		
	Cheuk, R., Shinn, P., Brooks, S., Buehler, E., Chao, Q., Johnson-Hopson, C., Khan, S., Kim, C., Altafi, H., Bei, B., Chin, C., Chiou, J., Choi, E., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Howing, B., Koo, T., Lam, B., Lee, J., Lenz, C., Li, J., Liu, A., Liu, J., Liu, S., Mukharsky, N., Nguyen, M., Palm, C., Pham, P., Sakano, H., Schwartz, J., Southwick, A., Thaveri, A., Toriumi, M., Vaysberg, M., Yu, G., Davis, R., Federspiel, N., Theologis, A. and Ecker, J.		
	Direct Submission		
TITLE JOURNAL	Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,		
	Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA		
	On Jun 28, 2000 this sequence version replaced gi:6007863. This submission of BAC F1504 is shorter by 1338 bases. The original BAC had a Tn10 transposon insertion (gb J01829.1 TRN10IS13) from E. coli located at the junction of bases 18229 and 18230 of this submission.		
FEATURES	Location/Qualifiers		
	1..158096		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3702"		
	/chromosome="1"		
	/clone="F1504"		
	528..659		
	/note="hypothetical protein"		
CDS	/codon_start=1		
	/evidence=not_experimental		
	/product="F1504.1"		
	/protein_id="AAF79342.1"		
	/db_xref="GI:8778334"		
	/translation="MLKANTFHLFMNNEVNQSTRMLKGYLKSRRITKEESNRDKCK"		
CDS	1377..1484		
	/note="hypothetical protein"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="F1504.2"		
	/protein_id="AAF79343.1"		
	/db_xref="GI:8778335"		
	/translation="MLRLPQLLASPKRTCSIASTAAATSCGDQTNMLIV"		
CDS	complement(join(2370..2492,2593..2748,2776..2868))		
	/note="similar to ubiquitin-conjugating enzyme E2D 2 gi 4507775"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="F1504.3"		
	/protein_id="AAF79390.1"		
	/db_xref="GI:8778382"		
	/translation="MIQPNGTPYASGIFRLDIEFFEDYFPKPPKVNFFVNINTFYD LFFNSTFKTPITYHSNIKSSGSICIDILKDKTSPSLTVEKLINKSHVLLSITVLLADPN PNDPLVPEIGQLFKNNRFQFD"		
CDS	complement(3065..3151)		
	/note="hypothetical protein"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="F1504.4"		
	/protein_id="AAF79389.1"		
	/db_xref="GI:8778381"		
	/translation="MSVKAKAMRRIETEFKAMSSKASLYSIG"		
CDS	complement(join(4475..4581,4720..5001,5076..5106,5210..5266))		
	/note="hypothetical protein"		
	/codon_start=1		
	/evidence=not_experimental		
	/product="F1504.5"		



```
/protein_id="AAF79388.1"
/db_xref="GI:8778380"
/translation="MTWRLTNLPQLTGHLEFTQDARNNSVFKKYTRNLFPPTSVPSTITT
WTKPTAPFVKCNFDAGFHINHSLSGTGGWIRDHQGSAAHAWGSSIVDHWNTPSEAEITKA
LLVAMQQGWTRKGYKMIQFEGDSHYLASSEFTNSIFQADTRNQPPWLANLLCKNLHS"
join(6536..6622,6671..6964,6989..7075)
/note="similar to phosphoenolpyruvate carboxykinase
(ATP)-like protein emb|CAB38935.1"
/codon_start=1
/evidence=not_experimental
/product="F1504.6"
/protein_id="AAF79344.1"
/db_xref="GI:8778336"
/translation="MISQFYIARYFLCIALEMFCYTEYLSVLLYEQAIEYKGSLLIT
SNGALTMLSGAKTDREISVLEMLRLRMSFGDYRIFILYFGLLIFVDFDFDKGLSN
IEIDEHTFMVVRERAVDYLNLEKILSHSHSGCMGKGDVAFELLGIAIF"
complement(join(9342..9431,9588..9662,9771..9872,
10106..10186,10497..10811))
/note="putative 50S ribosomal protein L21, chloroplast
precursor sp|P51412|RK21 ARATH; similar to ESTs
gb|T45781.1, and gb|AW004148.1"
/codon_start=1
/evidence=not_experimental
/product="F1504.7"
/protein_id="AAF79387.1"
/db_xref="GI:8778379"
/translation="MASSSATLSLCSFSAHCNVNSRRSSTILCSLSPKSLNLAKPLT
GFLSPSTASTSRTAFTVAPKFAESVVEAPETTDIEAVVSDVSEVTEEKAKREEIFA
VIMVGGRQYIVFPGRYLYTORLKDANVDDQIVLNKVLVGTKTHYIGKPVVNTATVH
AVVESQGLNDKVVFYKYPKKKRYRNIGHRQPNTRIRITGITGYEYPASPNNVAVGEV
NL"
complement(join(11530..11670,11756..11980,12064..12231,
12303..12418,12517..12669,12759..12902,13095..13225,
13537..14132))
/note="similar to calcium-dependent protein kinase SK5
sp|P28583|CDPK_SOYBN; similar to ESTs gb|AW728580.1,
gb|A1994248.1, and gb|A1997698.1"
/codon_start=1
/evidence=not_experimental
/product="F1504.8"
/protein_id="AAF79386.1"
/db_xref="GI:8778378"
/translation="METKPNRPRPSNTVLPYQTPRLRDHVLGKLLGQGQFGTTYLCT
EKSTSANYACKSIPIKRLVCREDYEDVWREIQIMHLSHPNVVRIKGTYESVVFVHI
VMEVCEGGELFDRIIVSKGHFSEREAVKLIKTLGVVEACHSLGVNHRDLKPFNLFDS
PKDDAKLKATDFGLSVFYKPLIFLFWLIDSLILQLVFWLFFVFSMNRKLKQSGFLIET
GLLCTFIANRNKVEGMYRFDLFGQYLYDVVGSPYYVABEVLLKKCYGPEIDVWSAGV
ILYLLSGVPPFWAETESGFIQILQGLKDFKSDPWPPTISEAAKDLIYKMLERSPKKR
ISAHEALCHPWIVDEQAAPDKPLDPAVLSRLKQFSQMNKIKKVALRVIAERLSEEEIG
GLKELFKMIDTNSGTITFEELKAGLKRVGSELMESETSIMDAADIDNSGTIDYGEF
LAATLHMNMKEREENLVAAFSYFDKDGSGYITIDELQSACTEFGLCDTPLDDMIKEID
LNDGKIDFSEFTAMMRKGDGGRSRTMMKNLNFNIADAFVDGEKSD"
join(17932..18090,19698..19920,20052..20233)
/note="similar to En/Spm-like transposon protein
gb|AAC14510.1"
/codon_start=1
/evidence=not_experimental
/product="F1504.9"
/protein_id="AAF79345.1"
/db_xref="GI:8778337"
/translation="MFKLSSAQKMLFWYAMLKAACNKYLKPSLTCGLPLISGTVTER
VLHDFITKVKIIGFMELISIDMFKPDPMTLTKVHPKLRSDRRYPHFKGFIGAMDGTH
VPAKVSGRDQORYWNRKNKCSNMILGKYFVDSGYGLRCGYLGPYRQSYRHPHFQNG
APPYNYKEFNRLHFSLSRVIERIFGV"
20230..20508
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F1504.10"
/protein_id="AAF79346.1"
/db_xref="GI:8778338"
/translation="MKRKMEDNASRARYDVRITRKLIVVATWALHNFVRKSNIPDDPFE
ANCEQEGNHQPSLNEEVEVQEDDQMIDSRQYMEGIRDDIAMNLWNNHR"
```

CDS

```
complement(join(21563..21662,21848..22122,22212..24551,
24643..24717,24801..24989,25089..25340,25419..25941,
26020..26227,26306..26741))
/note="unknown protein; similar to EST gb|N37451.1"
/codon_start=1
/evidence=not_experimental
/product="F1504.11"
/protein_id="AAF79385.1"
/db_xref="GI:8778377"
/translation="MEQSAAPQPTMPLPYLPSSVEASRDDLQCICTGMVIVPKPVGF
LCGISPLADNSFPASFTSALLPSQETLVTSPSSSSSSSSSSSSSYLAICLKLFFIC
CSVVTAPRYQMLPMETDLNLPFLTDFPDNVLPLAAAKSRITGDISKEANVITSNLSK
KCEALAVSGLVEYGDEIDVIAVDILKQIFKIPYSKARVSIQVQGVQTLVLNPGPDV
EEGKLIIRRHNQPKCTKNVDSLFLNFAMHSVRMEACDIPPTREHTEKRSSSSALP
AGENSHDNAPDDRDKPAGSKSQKQDGFICEKKSKKNKAGVEPVVRKNSQISEKIKS
SGSEKHSRGGSEFLRVLFQFHFNRMLLGSDDLFLSFNEKYVAVSLHLWDVSEKVT
LTWLEAWLDNVMAVPELAICYHENGIVQGYELLKTDDIFILKGISEDGTPAFHPHV
QONGLAVLRFLOSNCKEDPGAYWLYKSAGEDELQFLDLSISKNHSSSVHNDSASSPS
LIHSGRSDSMFSLGNLLYRVGHRLSLSVVPNDRNKCARELTQCLNCLDAPDHLVVRAY
AHEQFARLILNDSDESLTFESNGVQREVKITDLEEEALDPVTIADHENETVTFSEDK
FTEDHSVSNIVPLSVRPKLEANVSLCKELLHSDSPDHDTEGSAVNSSSDTSLDLGT
LCQTTTSPISSKLSAINHVSQAIKSLRWTRQLQSSQEQVAFHDILPDFSKSCSGDPDC
IEVDIRKWLPTSKLDRKLNVLILLGESYLSLGEAYKEDKQLHQAALNTVELACSIYG
SMPQKFEETLFVSSMKNLSLQSKFHERTQVEDLEAKSGSPDISVEELSTRLFWAKV
WMLVGDIVYQFHLKGQELSRRTKGTTHLKMQSEVVKVQRLKKKLTYSQNCASCS
LVNCSCKSDRASSGSSASSNGSSARTVPHSRKHNRLKQSNVASKVSRDVEDERVNF
KVENSKRKEEEDTSGETKGAVRLEQNESNSKETPGAKKGIFKYLKGSCTDDAESNLL
AALNCYEBETRRALQELPSNCSEFQSVLRKKGWVCNELGRNRLGSKELNKAEDAFADAI
VAFKEVCDHTNVILINCNLGHRRALAEEMVPKIEALHLEHRAFENAYQKALGTAKLEY
SKSLRYYMAAKTELSVATAEASSVDNLKVEVYTQANTYLRFGMLLANEDTTAAARE
QKNILENTHDSSSDGKSDLRKREVLSDAIREALALYESLGEIRKQEAFAFYLQLA
RYHKDCCLGLETERQSGRPKPESNVIQRAQYALLADRNWQKSMDFYGPENLPSMFL
TILTIERSALSTVSNFWQNLFWLLESALSRLLEGRHISKTYAESLRTEDPKLYTKFMAQ
LQMLVKRMLALSPLPSEGANKSQTCSRSGDSGKRLRELYKTSKSTNLCDLNAMHALIQV
ISCSLNVKCDSVKHKRVLPQGGSVNVLS"
join(27313..27574,27603..27721,27735..28416,28568..28584,
29183..29237,29327..29469,29559..29835,29998..30455,
30581..30652,30731..30820)
/note="hypothetical protein"
```

CDS

Query Match	8.8%;	Score	453.4;	DB	8;	Length	158096;
Best Local Similarity	81.6%;	Pred. No.	8.9e-62;				
Matches	551;	Conservative	0;	Mismatches	116;	Indels	8;
Gaps	2;						

Qy	1	TTTTTGGCGGAAAAATCTCGGGTTTACGTTTTTGGCGGAAAAATCTCGTGTTTACGTTTT	60
Db	52687	TTTTTGGCGGAAAAATACGGATTACGTTTTTGGCGGAAAAATCTCGGATTACGTTTT	52628
Qy	61	TGGCGGAAAAATCTCGGGTTTACGTTTTTGCAGAAAAATCACGGGTTTACTTTTTTTG	120
Db	52627	TGGCGGAAAAATCTCGGGTTTACATTTTTTGGCGGAAAAATCTCGGATTAC-GTTTTTGG	52569
Qy	121	CGGAAAAATCACGGATTATATGTTTTTGGTGGAAAAATTACGAGTTTACTTTTTCTCAAT	180
Db	52568	CGGAAAAATCACAGGTTTATGTTTTTGGTGGAAAAATTACGAGTTTACTTTTTCTCAAT	52509
Qy	181	TTTCATTGCTTGATATATTTAAGAAATTTGGAAAAATATTAAATTTATTAAATGGTTTAGA	240
Db	52508	TTTATCGCTTGATATATTTAAGAAATTTGGAAAAATATTAAATTTATTAAATGGTTTAGA	52449
Qy	241	TGIGTTGGTTAAACCTAAATTCGCATTGGTTTAGAGATTTTAGATTTGGTTTATTCAATTT	300
Db	52448	TGIGTTGGTTAAACTTAAATTGACATTGGTTTAGAGATTTTAGATTTGGTTTAAATGAATTT	52389
Qy	301	TACAAAAATTTGATGGGTTAAATTTGGATAAACCATGGAAACCATTAACCATTACAACCTAAC	360
Db	52388	TACAAAAATTTGATGGGTTAAATTTGGGTAAACCATTTGAACCATTAACCATTACAACCCAAC	52329
Qy	361	TCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACTCATCAACTCATTTGAG	420
Db	52328	TCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACTCATTAACCTCATTTGAG	52269
Qy	421	TCAAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACCCATGAAT	480

CDS

CDS

CDS

CDS

CDS



Db 52268 TCAAAATTTTCAACTCATTAGGGTTCAATGGGTTGAGTTGAGTTGATCCATGAATT 52209  
QY 481 TTGACCCATTTTGACACCCCTACATATGATC-----ATAAGTTAATAATCAAAAATTA 533  
Db 52208 TTGACCCATTTTGACACCCCTAGTTATATACCAATCCAAAAAGAAACGTTTGCAAATGAA 52149  
QY 534 CTATTGATAACTTTTAAACGGAATTGTTTCTGAAGTTTCATTTACGTTATTATATAAAA 593  
Db 52148 TCATCCACAAAACCTAATTGAGAAATCTGFACTAAACAGGTAGCCTTTAAATTGAGACGA 52089  
QY 594 AAACATAATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGCAAAAATATT 653  
Db 52088 CGAGAAAAAGAAAAAGAAAAAGAAATAGGTATTCCTTCTAGAAGTTTGAAACAAACCA 52029  
QY 654 ACTGTAATACTTTTT 668  
Db 52028 AACTTGTTTTTTTT 52014

RESULT 11  
AC011809  
LOCUS  
DEFINITION  
AC011809 Arabidopsis thaliana chromosome I BAC F6A14 genomic sequence,  
complete sequence.  
AC011809  
AC011809.2 GI:6579253  
VERSION  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 108767)  
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Unpublished  
2 (bases 1 to 108767)  
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,  
Chin,C., Chieu,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,  
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,  
Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,  
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,  
Theologis,A. and Davis,R.W.  
Direct Submission  
Submitted (15-OCT-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
3 (bases 1 to 108767)  
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,  
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,  
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,  
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.  
Direct Submission  
Submitted (15-DEC-1999) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
4 (bases 1 to 108767)  
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,  
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.  
and Davis,R.W.  
Direct Submission  
Submitted (22-JAN-2000) DNA Sequencing and Technology Center,  
Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
USA  
On Dec 15, 1999 this sequence version replaced gi:6041764.  
e-mail for correspondence: arab@sequence.stanford.edu

Genes with similarity to proteins in the databases are described as 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://gnomic.stanford.edu/~chris/GENSCANW.html), Fexa (V.Solovyev & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).  
FEATURES  
source  
1.108767  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3702"  
/chromosome="I"  
/clone="F6A14"  
4134.6402  
/gene="F6A14.1"  
join(4134.4785,4919.5062,5155.5307,5397.5512,5621.5788,5881.6111,6229.6402)  
/gene="F6A14.1"  
/note="calcium-dependent protein kinase 1, gb/S46283, the cloned cDNA for this protein starts at nucleotide 4160 of this BAC (gb|AC011809), 24 bp before the predicted start codon."  
/codon\_start=1  
/product="calcium-dependent protein kinase 1"  
/protein\_id="AAF27092.1"  
/db\_xref="GI:6730697"  
/translation="MGNCACVRPDSKESKPSKPKPNRDRKLNPFAGDFTTRSPAPI RVLKDVIPMSNQTIQSDKYILGRELGRGEFGITVLTCTDRETHEALACKSISKRLRTA VDIEDVRREVAIMSTLPEHPNVVKLKASYEDNENHVLVMELCGEGELFDRIVARGHYT ERAAAAVARTIAEVMVMCHSNGVMHRDLKPFNLFANKKENSPLKAIDFGLSVFPKPG DKFTEIVGSPYYMAPEVLKRDYGPVDVMSAGVIYILLCGVPPFWAETEQGVALLAIL RGVLDKRPDPWPQISESAKSLVKQMLDPDPTKRLTAQQVLAHPWIQNAKKAPNVPGLD IVRSLKQFSMMNRFKKVLRVIAEHLISIQEVEVIKNWFSMLDDDKGKITTYPELKAG LQKVGSQLGEPEIKMLMEVADVVDNGFLDYGEFVAVIHLQKIENDELFKLAFMFPDK DGSTYIELDELREALDELGEPPDASVLSDIMREVDVTDKGRINRYDEFVTMMKAGTDMR KASRQYSRERFKSLNIMKDGSLHLHDALTGQIVPV"  
complement(6629.9070)  
/gene="F6A14.2"  
complement(join(6629.7529,7607.8160,8258.8475,8980.9070))  
/gene="F6A14.2"  
/note="Similar to peptide transport proteins"  
/codon\_start=1  
/protein\_id="AAF27093.1"  
/db\_xref="GI:6730698"  
/translation="MEVEKTEKNIAEDDESKIIYRGWKVMPFIIGNTFEKLGIIVGSS SNLVIYLTTFNMKSIATAKVNIYGGTSNFGTIVA AFLCDSYFGRYKTLSEFAMIACF LGSVAMDLTAVIHELHPAQCAKEIGSVNCGPSICQIMFLAGAMVLLVIGAGGIRPCNL PFGADQFDPKTKEGKRGIESFFNWYFFTTFEAQMVSLTLIVVQSNVSWSIGLAIPAI LMLLGCIIFFAGSKLYVKVKASGSPHISITRVIVVAIKRRLKPVGPNELYNVIASDF KNSKLHTEQFRELDKSAIQOTDDKLNKDGSPDAWKLCMQQVEEVKCVIRVLPVWL SAALFYLAYIQQTTYTIFQSLQSDRRLPGSFQIPAGSYTVFLMLGMTIFTIYDRVL VPFLRKYTRDGGITQLQRVGAGLFLCITSMMVSAIVEQYRRKVALTKPTLGLAPRKG AISSMSGMWLIPQLVLMGIADALAGVQMEFYKQFPENMRSFAGSLYCGIGLASYL STFLLSAVHDTTEGSGGSLWLPEDLNKGRLEYFFLVAGMNTLNLAYFLLVSHWYRYK DVVAKDKMDKTSAEFDKVSV"  
complement(10694.14124)  
/gene="F6A14.3"  
complement(join(10694.10741,10837.10929,11012.11109,11244.11328,11532.11642,11809.11866,11962.12080,12177.12245,12346.12420,12498.12644,12724.12840,12920.13037,13131.13366,13960.14124))  
/gene="F6A14.3"  
/note="Putative isochorismate synthase"  
/codon\_start=1  
/protein\_id="AAF27094.1"  
/db\_xref="GI:6730699"



AUTHORS Lin,X. and Kaul,S.  
TITLE Direct Submission  
JOURNAL Submitted (28-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
REFERENCE 3 (bases 1 to 80139)  
AUTHORS Town,C.D. and Kaul,S.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
COMMENT On Jan 19, 2001 this sequence version replaced gi:12280901.  
Address all correspondence to:at@tigr.org

BAC clone T8E3 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.  
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/softlab/glimmerm\_htm/glimmerm.html, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).

FEATURES  
source  
1. .80139  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="T8E3"  
complement(894. .1959)  
/gene="T8E3.14"  
complement(join(<894. .1112,1254. .1457,1801. .>1959))  
/gene="T8E3.14"  
complement(join(894. .1112,1254. .1457,1801. .1959))  
/gene="T8E3.14"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG51277.1"  
/db\_xref="GI:12322548"  
/translation="MAHIMWKVLTNTNDKGAKRRELTPLRNTRQKYGERQDTNASWLSDDRTIKPLPSCGFRSMLPLNELLDVGGFMVNGEVKIVAEGVLEVVVRKSDVLVE TSLVMESIDVNGFQVLPSQVESVKSLEKHPDIAKFRPKNPHLRTTSLNSLLSLTHEI LCHCQSPEELSIDDLANAYSTLICLTKAGFKLD"  
complement(3047. .3695)  
/gene="T8E3.3"  
/note="predicted by genscan"  
complement(join(<3047. .3397,3519. .>3695))  
/gene="T8E3.3"  
complement(join(3047. .3397,3519. .3695))  
/gene="T8E3.3"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG51268.1"  
/db\_xref="GI:12322539"  
/translation="MVPLTKLLDKNKGFMVNGEVKIVAEGVLEVVGKSDVLEETSLV MESVMVNGFQVLPSQVEFVKSLEKHPDIESKHLHKNPHRITYLNSLLTLELSQS PENISNDDLANAYSTLSYVTKAGFKLDWLEKDLKDVGETRIQIEIEEELKDIKQKQVEM KHKCVDMEEALLEFLL"  
complement(5787. .6988)

gene  
/gene="T8E3.4"  
/note="predicted by genemark hmm"  
complement(join(<5787. .6092,6218. .6444,6618. .6784, .6882. .>6988))  
/gene="T8E3.4"  
complement(join(5787. .6092,6218. .6444,6618. .6784, .6882. .6988))  
/gene="T8E3.4"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG51271.1"  
/db\_xref="GI:12322542"  
/translation="MEDQYEKKITWTIKNFSVQSAIDSDIFVVGDSKWHLVAYPKG NGBSTNKLCLSYLNVADEFQSLPNGMKRHIYRLTVVNQMSSEKLSQEVIQGWFYKNPH ISGFQTMPLSKLLDKNGFLVNGDVKIIVVEVLEVVGKSDVLEETLLVHESIDING FQVLPSQVESVNNLFEKHPDIVSEFRSKNPLMRTTYLNDLLCLTEILCESSEELSTGD MANAYSTLTCLKKAGFKLDWLEKKLKEVCEARVQEIDDEWKDLTDLKEKF"  
complement(7821. .9077)  
/gene="T8E3.21"  
complement(join(<7821. .8165,8304. .8521,8706. .8872, .8971. .>9077))  
/gene="T8E3.21"  
complement(join(7821. .8165,8304. .8521,8706. .8872, .8971. .9077))  
/gene="T8E3.21"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG51272.1"  
/db\_xref="GI:12322543"  
/translation="MADQYEKRITWTIKNFSSLOSHAIYFDIFVVGDTKWHLLAYPKG YGDSINKCLSLFLGVPDPDDLPSGWKRHIIYRLTVVNQMSSEKLSQEVARGGFYPRSL TFGSQVMLPLTELYGGFLVSGQKIVAEGVLEVVGKSDVLEETLLVNGGINVNGFQV LPSQVESVNNLFKNHPDIASNFRLENTHLRTTYLNLNLCLTELLCQSPHKLSNVDLAN AHCTLTCTVKAGFKLDWLEKKLKEVGKTRMQQLEQNLKDKLESWCSSDDEDLRSV KT"  
complement(9690. .11941)  
/gene="T8E3.6"  
complement(join(<9690. .10133,10250. .10411,10547. .10729, .10818. .10958,11116. .11364,11460. .11599,11680. .>11941))  
/gene="T8E3.6"  
complement(join(9690. .10133,10250. .10411,10547. .10729, .10818. .10958,11116. .11364,11460. .11599,11680. .11941))  
/gene="T8E3.6"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAG51275.1"  
/db\_xref="GI:12322546"  
/translation="WATAAIIISVKPPASFYCGGNRRDGGRGITIAVRSRDERFHERIV GLSASTVLFIGLSLRFVCSPASARILPPIVVTVNEINTETVAGEENLNEVEDEKLE ADFEAYKSKVYSLTVPLKLVALRGSVPPSWIKFEMSSQGRVRLKTQFRGNLEEIFFD LSKPSKNGKKGAASTAAADMISIGDSSLFAIKEKLIIEPMKGIEDQDWYKGLSDKWKS QIYLRRNYAGEKADGETWAVPYRWGTMVIAYKSKFQNYKLAPIEDWADLWRPELAG RIAMVNSPREVVGAVLKYMRASYNNTDLSQVPGRLAVEKNLASLMKQIRLFDNNY LKAFNVGDVWTVVGSSSDVIPVAKRMSNVTIVPKSGATLWADLWAIPAVSDSCKEAE QRGGRVGPSPLINQWIEFCLQPARSLPFTREVTPGASPSALDGPLVTEPEKTKDRT KLDTNLVTGVPPEILSKCEFLPEPLPEATLSEYRLLIETVRKQSQRPGLVEKLDIVS IKVRGFRAKLDSEMNKNI"  
12959. .16475  
/gene="T8E3.2"  
/note="contains Pfam profile: PF00069: Eukaryotic protein kinase domain"  
join(<12959. .13055,13146. .13281,13357. .13428,13506. .13577, .13677. .13748,14043. .14114,14198. .14269,14497. .14585, .14679. .14837,14944. .15276,15355. .15600,15677. .15810, .15894. .16475)  
/gene="T8E3.2"  
join(12959. .13055,13146. .13281,13357. .13428,13506. .13577, .13677. .13748,14043. .14114,14198. .14269,14497. .14585, .14679. .14837,14944. .15276,15355. .15600,15677. .15810, .15894. .16112)  
/gene="T8E3.2"  
/codon\_start=1  
/product="protein kinase, putative"



/protein\_id="AAG51266.1"  
/db\_xref="GI:12322537"  
/translation="MGICEMKSCSWLLISLLCSLSNESQAISPDPGEALLSFRNAV  
RSDSFIHQWRPDPDPFCNWNNGVTCDAKTKRVTNLNLYHKIMGPLPPDIGKLDHLRL  
MLHNNALYGAIPTALGNCTALEEIHQSNYFTGPIPAEMGDLPGIQKLDMSNTLSGP  
IPASLQKLKLSNFNVSNNFLVQIPSDGVLSGFSKNSFIGNLNLCKGKHVDVVCQDD  
GNPSSHQSQONQKNKSKLLISASATVGALLVALMCFWGCFLYKKGKVEIKSLAK  
DVGGASIVMFHGDLPYSSKDIIKLEMLNEEHIIGCGFGTVVKLAMDDGKVFALKR  
ILKLNEDRDPFFERELEILGSIKHYLVNLRGYCNSPTSKLLLYDLPGGSLDEALHE  
RGEQLDWDNRVNIIGAAKGLSYLHHDSPRIIHRDIKSSNILLDGNLEARVSDFGLA  
KLEDEESHITIVAGTFGYLAPEYMQSGRATEKTDVYSFGVLVLEVLGSKRPTDASF  
IEKGLNVVGLKFLISEKRPRDIVDPNCEGMQMESLDALLSIATQCVSPSPERPPTH  
RVVQLLESEVMTCPSEFYDSSD"  
complement(16532. .18333)  
/gene="T8E3.8"  
/note="contains multiple PPR domains: PF01535: PPR repeat"  
complement(<16532. .>18333)  
/gene="T8E3.8"  
complement(16621. .18333)  
/gene="T8E3.8"  
/codon\_start=1  
/product="PPR-repeat protein"  
/protein\_id="AAG51260.1"  
/db\_xref="GI:12322531"  
/translation="MNMSLIQTPSLLMYNKMCLKSLADGKSFYKVLALFGELRGQGLYP  
DNFTLPVVLKSIIGRLRKVIEGKHYAVKAGLEFDSYVSNLGMGYASLGKIEITHK  
VFDEMPQORDVVSNGLISSYVGNRPFEDAIGVFKRMSQESNLKFDDEGTIVSTLSACA  
LKNLEIGERIYRFVVFTEFMSVRIGNALVDMFCGCLDKARAVFDSMRDKNKCWTS  
MVFGVSTGRIDEARVLFRSPVKDVVLWTAMNGYVQFNRPDEALELFRCMQTAGR  
PNFVILSLTGTCAQTGALEQGWIKHYINENRVTVDKVGTALVDVYAKCGCIEFAL  
EVFEIKERDTSWTSLIYGLANWMSGRALDLYEMENVGVRDLDAITFVAVLTACNH  
GGFVABGRKIFHSMTERHNVPKSEHCSCLIDLRCAGLLDEAEELIDKMRGESDETL  
VPVYCSLLSAARNYGNVKIAERVAEKLKVEVSDSSAHTLLASVYASANRWEDVTNVR  
RKMKDLGIRKFPFGCSSIEIDGVGHEFIVGDDLLSHPKMDEINSLHQTNNMLDLEHK  
EIDS"  
complement(18475. .21274)

Query Match	8.7%;	Score	449;	DB	8;	Length	80139;
Best Local Similarity	93.5%;	Pred.	No. 5e-61;				
Matches	491;	Conservative	0;	Mismatches	30;	Indels	4;
Gaps	2;						

QY	1	TTTTTGGCGGAAAAATCTCGGGTTTACGTTTTCGCGGAAAAATCTCGTGTTTACGTTT	60
Db	49768	TTTTTGGCGGAAAAATTATGGATTACATTTTGGCGGAAAAATCTCGGTTTACGTTT	49827
QY	61	TGGCGGAAAAATCTCGGGTTTACGTTTTCGAGAAAAATCACGGTTTACTTTTTTTG	120
Db	49828	TGGCGGAAAAATCTCGGGTTTACGTTTTCGAGAAAAATCACGGTTTACTTTTTTTG	49887
QY	121	CGGAAAAATCACGGATTTATGTTTTTGGTGGAAAAATACGAGTTTACTTTTTCTCAAT	180
Db	49888	CGGAAAAATCACGGATTTATG-TTTTTGGTGGAAAAATACGAGTTTACTTTTTCTCAAT	49946
QY	181	TTCAATTGCTGTATATTTAAGAAATTTGGAATAATTAATTTATTAAATTTGTTTAGA	240
Db	49947	TTCAATCTCTGTATATTTAAGAAATTTGGAATAATTAATTTATTAAATTTGTTTAGA	50006
QY	241	TGTGTTGGTTAAACCTAAATTGGCAATTCGTTTAGAGATTTTGGTTTATTCAATTT	300
Db	50007	TGTGTTGGTTAAACCTAAATTGGCAATTCGTTTAGAGATTTTGGTTTATTCAATTT	50066
QY	301	TACAAAATTTGATGGGTTAATTGGATAAACCATGGAAACCATTAACCATTAACCTAAC	360
Db	50067	TACAAAATTTGATGGGTTAATTGGGTAACCATTTGAAACCATTAATCATTAACACCA	50126
QY	361	TCATTTTACTCATCAACCAATGACTCATCAACTCATTTGACTCATCAACTCATTTGAG	420
Db	50127	TCATTTTACTCATCAACCAATTGAC--TCAACTCATTTGACCATCAACTCATTTAAG	50183
QY	421	TCAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACCCATGA	480
Db	50184	TCAAAAATTTTCAATTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACCCATGA	50243

QY	481	TTGACCCATTTTGACACCCCTACATATGATCATATAAGTTAATAATC	525
Db	50244	TTGACCCATTTTGACACCCCTAACTATAGTATGATATTTATTAAC	50288
RESULT	13		
AC006551			
LOCUS	AC006551	107200 bp	DNA linear
DEFINITION	Arabidopsis thaliana chromosome I BAC F12K8 genomic sequence, complete sequence.		PLN 11-JUN-2001
ACCESSION	AC006551		
VERSION	AC006551.6	GI:6056182	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 107200)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 107200)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (10-FEB-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	3 (bases 1 to 107200)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P., Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-OCT-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	4 (bases 1 to 107200)		
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Araujo,R., Huizar,L., Rowley,D., Brooks,S., Buehler,E., Chao,Q., Dunn,P., Gonzalez,A., Khan,S., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vyotskaia,V., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-DEC-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Oct 16, 1999 this sequence version replaced gi:5881526. Bases 1-12905 of clone F12K8 overlap with bases 70259-83163 of TAMU BAC clone T22J18, gb AC003979.		
FEATURES	Location/Qualifiers		
source	1. .107200		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3702"		
	/chromosome="I"		
	/clone="F12K8"		
gene	complement(816. .1713)		
	/gene="F12K8.1"		
CDS	complement(join(816. .1326,1407. .1536,1614. .1713))		
	/gene="F12K8.1"		
	/note="Location of ests 250F14T7 (gb AA598046), 250E15T7 (gb AA598042), 110A14T7 (gb T42000 5263), and 1B4T7P (T04118). 58% identical to DNA-binding protein		



```

[Arabidopsis thaliana] (gb|U26936).
/codon_start=1
/product="Putative DNA binding protein"
/protein_id="AAF18515.1"
/db_xref="GI:6587826"
/translation="MNKGAWTKEEDQLLDYIRKHGEGCWRSLPRAAGLQRCGKSCRL
RWNYLRPDLKRGNFTEEEDELIIKHLSLLGNKWSLIAGRLPGRTDNEIKYWNTHIK
RKLLSRGIDENSHRLINESVSPSLQNDVETIHLDFSQPVKPEPVREEIGMVNNE
SSGTTSEKDYGNEDDWLNLELSVGPSRYESTRKVSVDSEAESTRWGSELFGAHES
DAVCLCCRIGLFRNESCRNCRVSDVRTH"
complement(3868. .4548)
/gene="F12K8.2"
complement(join(3868. .3955,4062. .4108,4259. .4318,
4411. .4548))
/gene="F12K8.2"
/note="Unknown protein; Location of ests 89E10XP 3'
(gb|A1100452), 207C19T7 (gb|N37611), and 89E10T7
(gb|T20896)"
/codon_start=1
/protein_id="AAF18516.1"
/db_xref="GI:6587827"
/translation="MLKFPLCCSELSPKILAFESQLVSGRRDTIKVSAGKIGNFSL
GSIFKSCETCGAKGAIECPGCKGTGRNKNGNMFERWKCFDCQCGFMKSCPKCGKGL
TPEQRGR"
5248. .10146
/gene="F12K8.3"
join(5248. .5379,5566. .5770,5860. .6025,6152. .6374,
.6452. .6528,6752. .6939,7064. .7161,7227. .7384,7457. .7591,
7713. .8112,8641. .8726,8916. .9128,9317. .9502,9690. .9801,
9895. .10146)
/gene="F12K8.3"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAF18517.1"
/db_xref="GI:6587828"
/translation="MAKSENSTTSTFFSFANKIQPSNDAESDPDSYALEKFLYETRA
RFYLVSDRNKRFVRVLKIDRSESELNISEDPPVYSPQETKSLLQRIAEGNRATGGL
AFVAKVYGIAGCAKEMESYYLVLVTKRQIGCICGHAIYAIDESQMISVPHATIOQDV
ANSKTELRYKKLLSSVDLTCKDFPYSTYPIMQSLQKNVLSGEEGMPYDNI FVWNSYL
TQPIRSRCNNTIWTALVHGHEQIRLSIYGRDFSVLVSRSRHFAGTRYLKRGVND
RGRVANDVETEQLVLDDEAGSCGKMSSVQMRGSIPLFWSQEASRFSPKPDIFLQRY
DPTVESTKMHFEDLVNRYGNPIVLNLIKTHGNIQTVEKRPREMVLRRREFANAVGY
LNSIFREENHLKFIHWDFFHKFAKSKSANVLAVLGAVASEALDLTGLYFSGPKPIVKKK
ASQLSHANTARDKYCREPSRLDRAYSaelSRGESANDILSALANREKEMKLTQOKKD
EGTNSAPRYQSGVLRTNCIDLDRTNVAQYAYGLAALGFQLHAMGLSDTPKIDPDSS
IAAALMDMYQSMGDALAQYGGSAAHNTEGKPALWELDDSYLYHVGIGDIFPDIGV
QSIAPKPMSGIGVNLAPVAFRDPFSRKKLTsFDKLIETCSSIKNVRLCSETDQRPGG
NWTSGVAPDAAEIQLKSPNWLFGSRKPEESSATSGADDSEKGVSTSTERVNDFCNL
DWLSKDRHQGDIQORYLSTITNEANGWYGGTLLGDQDENSEIYRHYAQFCQCFAME
PFENDHEFEQNFAEVLRMTIDVMDIEEETEEMESDFNEYTQIGSDLGIIIPMOCKHFA
SDPCWLARWLVGDDKVPKVI"
complement(10468. .13557)
/gene="F12K8.4"
complement(10468. .13557)
/gene="F12K8.4"
/note="Last 369 amino acids are 61% identical to
phosphoribosylanthranilate transferase [Pisum sativum]
(gb|BAA13032)."
/codon_start=1
/product="Highly similar to phosphoribosylanthranilate
transferase"
/protein_id="AAF18518.1"
/db_xref="GI:6587829"
/translation="MNKLVEIVDASDLMPKDGQGSAPFVEVEFDEQRQRTQTRFKD
LNQOWNEKLVFNVDGLKRLNKNVTVDVYDDRRDNQPGKFLGRVKIAGAVVLSESES
GVQRYPLDKRGLFSNIKGDIALRIYAAPIDGDFVSPDPDFAEKVMKEDKRFESEQFQ
FONQONQNHVEQFEDEINNMETLKPTKKKEKESRTPHISIGAHAGGGGAPMWSQAKQ
AYPPPNQPEFRSDMRAPGPTGAVMQMPPPROQNPPEFQLIETSPPLAARMRQSYVY
RSSGDKTSSYDYLVEQMHYLYSVVKARDLPVMDVSGSLDPYVEVKLGNKGLTKHLE
KNSNPiWKQIFAFSKERLOSLLLEVTKDKDLTFDVGVRVHIDLTEVPLRVPDPS
LAPQWYRLEDKKGMKTNREIMLAVMVGTOADESPDAWSDAHRVSHSNLSNTRSKV
YFSPKLYLRIHVMEAQDLVPDCKGRVPDAI VKIQAGNQMRATRTPQMRTNWPQWHEE
LMFVVSPEFEDMIVSVDDRIGPGKDEILGRVFIpVRDVPVRQEVGKMPDPRWFNLQR
```

```

HSMSEBENEKREKESKILLRVCIAGYHVLDESTHFSSDLQPSKKHLRKPISGIL
ELGILSARNLMPMKGDGRMTDPYVAKYGNKWVRTRTLDDALAPKWNEQYTWEVHDP
CTVITIGVFDNSHVNDGGFDKQRIQKVRSLTLETRVYTHFYPLLVLTGGGLKKN
GELQALRYTCTGFVNMAQYGRPLLPKMHIYQIPIVRHIDLLRHQAQIIVATRLSRS
EPLRREVEYMLDVDYHMFSLRSKANFSRIMSLSSVTLVCKWFNDICTWRNPITIT
CLVHVLFLILVCYPELILPTVFYLFVIGMWNRYRPRPHPHMDARVSDADNAHPDEL
DEEFTPTPTSRPADIVRMRYDRLSVGGRVQTVVGDLATQGERIQALLSWRDPATAL
FIVEFALIWAFFIYVTFQVIAIIGLFMLRHPFRSRMPSVPANFFFKRLPAKSDMLL"
18848. .20727
/gene="F12K8.5"
join(18848. .19732,20455. .20727)
/gene="F12K8.5"
/note="36% identical to seed maturation protein PM27
[Glycine max] (gb|AAD30426)."
```

gene	/codon_start=1
CDS	/product="Similar to seed maturation protein PM27"
	/protein_id="AAF18519.1"
	/db_xref="GI:6587830"
	/translation="MTKLLALSILVLSALFTVGHCTGSEDDIIVRDGHRVVVVEYDRDG
	ETNTRVLISPPGKEVDQKSRDEGEVFGNEKRETAASSLPBEEERHHATPGELICDAIG
	KCKHKLGTVLGRVKDRDTASDLSDETPEMTVAREALEVEKVSWKAREARGKVNERATK
	KAHRVQKVLQVIAVRGIGTVVATALGLTKIGSVVGIVGIAAAYCMCVWTVFVSGYV
	LASVLGEQQQFVVQSKMYPVYFKAVSVGILVGLLGHVTCRRRKVFTDAVDMWQSVNLL
	SSLIMVEANASFVYTRATKAMFELIKAEKEDGRGFDTSQSSSESAGRTRGKKKYTE
	KTDEDVVKQRLTKLSERMRKLNAYSSRLNLLTLMSLTWHFVYLYGYRLSLTC"
gene	21976. .22641
CDS	/gene="F12K8.6"
	join(21976. .22094,22206. .22641)
	/gene="F12K8.6"
	/note="Hypothetical protein"
	/codon_start=1
	/protein_id="AAF18539.1"
	/db_xref="GI:6587850"
	/translation="MMGLFVLENFPGFKLIGGKISGLRFWRKITDLRFWRENLGFTFL
	AEKSRVYVFDGKIMGLRFGKKNKGFTFLAEKSRVYVSGKITGLRFWRENHGVTFLAE
	KSRVYVFGGKITGLRFWRENYGFTFLAGKLWIYVFGKITGLRFWRENHGFMEFGGKI
	TSLFFLNLCIACIFKKFGKILILLN"
gene	complement(24033. .26186)
CDS	/gene="F12K8.7"
	complement(join(24033. .24315,25922. .25990,26161. .26186))
	/gene="F12K8.7"
	/note="Unknown protein; Location of ests F13558
	(gb F13558) and Z37169 (gb Z37169)"
	/codon_start=1
	/protein_id="AAF18520.1"
	/db_xref="GI:6587831"
	/translation="MTMAPKRCFKTNESTKSVRVACRSLSRSPSVRRLSELPEVKQTKY
	MMQKDLNMKMIQDAEKKLEKEKMHTRAMKLGLMAGSNDLITDTCSEELARAADVVD
	KKLKAIRERIKAVEAGAPIIKRD"
gene	29462. .31415
CDS	/gene="F12K8.8"
	join(29462. .29585,29667. .29884,29981. .30540,30620. .31415)
	/gene="F12K8.8"
	/note="41% identical to LeOPT1 [Lycopersicon esculentum]
	(gb AF016713)."
	/codon_start=1
	/product="Similar to LeOPT1 [Lycopersicon esculentum]"
	/protein_id="AAF18521.1"
	/db_xref="GI:6587832"
	/translation="MKIPEEEVALLEDYVSDVDHRGFPAGKSSSTGGWPSAWFIIGVE
	VAERFAYFGIACNLITYLTGFLGQSTAKAAVNVNTWSGTASILPILGAFVADAYLGRY

Query Match	8.7%;	Score	449;	DB	8;	Length	107200;
Best Local Similarity	80.6%;	Pred.	No. 4.7e-61;				
Matches	549;	Conservative	0;	Mismatches	130;	Indels	2; Gaps 2;
Qy	1	TTTTTGGCGGAAAAATCTCGGTTTACGTTTTCGCGGAAAAATCTCGTGTTTACGTTT	60				
Db	22409	TTTTTGGTGGAAAAATTACGGGTTTACGTTTTTTTGGCGGAAAAATTACGGGTTTACGTTT	22468				
Qy	61	TGGCGGAAAAATCTCGGTTTACGTTTTTTTTCGCGAAGAAAAATCACGGGTTTACTTTTTTG	120				
Db	22469	TGGCGGAAAAATTATGGAATTACGTTTTTTTGGCGGAAAAATCACGGGTTTAC-GTTTTTGG	22527				



```
/product="hypothetical protein"
/protein_id="AAG60155.1"
/db_xref="GI:12597845"
/translation="MAGISRVLGCFRTLMSKAATVEAVTVAGEGKGIFKTRPVSQOLA
SFAGRELTRGSALKKVVYVKLHNLQNPA NKKEIHCDDKLKTI FDGKDKVGITEIMK
LLSPHPKSV"
gene
7467. .14213
/gene="F27M3_5"
/note="contains similarity to cytochrome c"
join(<7467. .7707,8400. .>8689)
mRNA
/gene="F27M3_5"
join(7467. .7707,8400. .8689)
CDS
/gene="F27M3_5"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG60137.1"
/db_xref="GI:12597827"
/translation="MGKDDHHEQDKRFFSHNNHHGHGYPGAYPPPPQGYPPPGGYP
POGYPPPHGYPPAAAYPPPPGAYPPAGYPPGPGPGFGGKGGLIAGATAAAAAMG
GHHAGHHGGYGHGHGKYKRGFPGGKYKRKGHSMFGGGKYKRKGKHGMFGGKRSHGM
FGGKRGKGMFGRKKWK"
mRNA
join(<9659. .9811,9913. .10121,10869. .10935,11021. .11113,
11276. .11419,11508. .11596,11758. .11863,11957. .12044,
12129. .12247,12334. .12495,12603. .12903,12990. .13099,
13178. .13358,13493. .13622,13826. .>14213)
/gene="F27M3_5"
join(9659. .9811,9913. .10121,10869. .10935,11021. .11113,
11276. .11419,11508. .11596,11758. .11863,11957. .12044,
12129. .12247,12334. .12495,12603. .12903,12990. .13099,
13178. .13358,13493. .13622,13826. .14213)
/gene="F27M3_5"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG60136.1"
/db_xref="GI:12597826"
/translation="MVSLSFILCCVLVSSCAYATIVSHDGRAITIDGHRVLLSGSIH
YPRSTEMWPDLIKKGEGSLDAIETYVFWNAHEPTRROYDFSGLDLIRFLKTIQNE
GMVGLRIGPYVCAEWNYGGFPVWLHNMPGMEFRITNTAFNEMQNFTMI VEMVKE
KLPASQGGPIILAQIENEYGNVIGSYGEAGKAYIQWCANMANSLDVGPWIMCQDDA
PQMLNTCNGYYCDNFSNNPTPKMTWNTWTGWYKNWGGKDPHRTTEDVAFVAFRF
QKEGTQNYMYHGGTNEDRTAGGPYITITDYDYAPLDEFGNLNQPKYGLKQLHDVL
HANEKTLTYGNISTVDFGNLTATVYQTEEGSCFIGNVNETSDAKINFQGTSDVPA
WSVSI LPDCKTETYNTAKINTQTSVMVKKANEAE NEPSTLKS WRPENIDSVLLKGKG
ESPMRQLFDQKWSNDESDYLWYMTTNLKEQDPVLGKNMSLRINSTAHLHAFVNGQ
HIGNRVENGKFHYVEQDAKPNGANVITLLSIITVGLPNYGAFPFENFSAGITGPVFI
IGRNGDETIVKDLSTHKSYKTGLSGFENQLFSPTWSAPLGSEPVVVDLILGLK
GTAWINGNNIGRYWPAFLSDIDGCSAEYHVPRSFNLSEGDNTLVLFEEIGGNPSLVNF
QTIGVGCVCANVYEKNVLELSCNGKPISAIKFASFGNPGD CGSFEKGTCEASNAAA
ILTQECVGKEKCSIDVSEDFGAECGALAKRLAVEAIC"
gene
complement(15582. .19591)
/gene="F27M3_7"
/note="similar to epsilon adaptin GI:5689377 from (Homo
sapiens)"
complement(join(<15582. .16556,16628. .16771,16854. .16931,
17301. .17417,17501. .17557,17793. .17924,18017. .18214,
18313. .18423,18506. .18706,18803. .>19591))
/gene="F27M3_7"
complement(join(15582. .16556,16628. .16771,16854. .16931,
17301. .17417,17501. .17557,17793. .17924,18017. .18214,
18313. .18423,18506. .18706,18803. .19591))
/gene="F27M3_7"
/codon_start=1
/product="epsilonon-adaptin, putative"
/protein_id="AAG60138.1"
/db_xref="GI:12597828"
/translation="MEQLKTIKRELAMGSGQGFGQSKBFLLDKVKSIGEARSKAEDRI
VLSEVDILKRLLLEPDI PKRMKEYI IRLVIYIMLGH DASFGYIYAVKMTHDNL LK
RTGYLAVTLFLNEDHDLIILIVNTIQDLRSDNYLVVCAALN AICRLINEETIPAVLP
QVVELLNHQKEAVRKA IMA LHRFHRKSPSSVSHLVSNFRKRLCDNDPGVMGATLCPL
FDLISEDVNSYKDLVSSFVSI LKQVTERRLPKSYD YHQMPAPFIQIMALLGSGDKNAS
DIMSVMVGLDLFRKCDSSNTIGNAILYECIRCISCLPNKLEAADAISKFLKSDSH
NLKXMGIDGLRLIKISPDIAEQHLAVIDCLEDDPDTLKRKTFELLYKMTKSSNVEV
IVDRMIDYMISINDNHYKTIASRCVELAEQFAPS NQWFIQIMNKVFEHAGDLVNIKV
```

```
AHNLMLRI AEGFGEDDDADSKLRLSAVESYQLI SEPKLP SLFLQVISWLGEYCTA
DGKYSASYISGKLC DVADAYSDET VKGYAVSALMKIYAF EIASGRKVDVLP ECOSLI
EELLASHSTD LQQRAYELQALIALDARAVETILP DASCEDIEVDKDL SFLNGYIQQA
IESGAQPYI SERESGMFETD YHPQDHVEP THALRFEAYELPKPSVP PQASNELVP
VPEPSYSESHQPISTLSVSERESSEI KLRLDGVKQWGRPSYQSTTAA SSTTPOAAN
GISTHSDAGV GSSSKPRSSYEPKPEIDPEKORLAASLFGSSSR TDKRSSSGGHKP
AKGTANKTATVPKENQTPVQPPFDLLDFGEPTATTAMD PPFKELEGLMDS SSGDGS
SDVMGLYSDAAPVTTTTSVDLSLSELS DSSKGN SRTYQPQTSKGPNTKEALEKDALVR
QMGVNPTSQNP TLFKDLLG"
gene
complement(22142. .23329)
/gene="F27M3_8"
/note="predicted by genemark hmm"
complement(join(<22142. .22194,22558. .22775,22906. .23126,
23204. .>23329))
mRNA
/gene="F27M3_8"
complement(join(22142. .22194,22558. .22775,22906. .23126,
23204. .23329))
CDS
/gene="F27M3_8"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAG60140.1"
/db_xref="GI:12597830"
/translation="MEIQQDNDRDPKPSFLPFMFIFL FGLAAFFLCLSAEFQKAKAL
LRAQVFLKGDLKWDGESC YLPENRAFGLIAALVCVSVAQIVGNVVICRGFTKDKT
RTTIFCIILLF SVWNFAVATLISVGASMNREQIYKGWLNR ECVLVKDG VFAASGF
Query Match 8.7%; Score 449; DB 8; Length 110664;
Best Local Similarity 93.5%; Pred. No. 4.7e-61;
Matches 491; Conservative 0; Mismatches 30; Indels 4; Gaps 2;
Qy 1 TTTTGGCGGAAAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGTGTTTACGTTTT 60
Db 92326 TTTTGGCGGAAAAATATGGATTTACATTTTGGCGGAAAAATCTCGGTTTACGTTTT 92267
Qy 61 TGGCGGAAAAATCTCGGTTTACGTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTG 120
Db 92266 TGGCGGAAAAATCTCGGTTTACGTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTG 92207
Qy 121 CGGAAAAATCACGGATTATGTTTTTGGTGAAAAAATTACGAGTTTACTTTTCTCAAT 180
Db 92206 CGGAAAAATCACGGATTATG-TTTTTGGTGAAAAAATTACGAGTTTACTTTTCTCAAT 92148
Qy 181 TTCATTGCTTGTA TATTTAAGAAATTTGAAAAAATATTAA TTTTATTAAATGGTTTAGA 240
Db 92147 TTCATCTCTTGTA TATTTAAGAAATTTGAAAAAATATTAA TTTTATTAAATGGTTTAGA 92088
Qy 241 TGTGTGGTTAAACCTAAATGGCATTTGGTTTAGAGATTTTAGATTGGTTTATTCAATT 300
Db 92087 TGTGTGGTTAAACCTAAATGGCATTTGGTTTAGAGATTTTAGATTGGTTTATTCAATT 92028
Qy 301 TACAAAAATTTGATGGGTTAATTTGGATAAAACCAATGGA AACCATTAACCATTAACAACCTAAC 360
Db 92027 TACAAAAATTTGATGGGTTAATTTGGTAAACCATTTGA AACCATTAATCATTACAACCCAC 91968
Qy 361 TCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACTCATCAACTCATTTGAG 420
Db 91967 TCATTTTACTCATCAAAACCAATTGAC--TCAACTCATTTGACCCATCAACTCATTTAAG 91911
Qy 421 TCAAAAAATTTCAACTCATTAGGGTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAATT 480
Db 91910 TCAAAAAATTTCAATTCAATTAGGGTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 91851
Qy 481 TTGACCCATTTTGACACCCCTACATATGATCATATAAGTAAATC 525
Db 91850 TTGACCCATTTTGACACCCCTAACTATAGTATGATATTATTAAAC 91806
```

```
RESULT 15
AC022522/c
LOCUS AC022522 88643 bp DNA linear PLN 12-SEP-2000
DEFINITION Arabidopsis thaliana chromosome I BAC T28K15 genomic sequence,
complete sequence.
ACCESSION AC022522
```



VERSION	AC022522.2	GI:6970637	
KEYWORDS	HTG.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1	(bases 1 to 88643)	
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miranda,M., Brooks,S., Buehler,E., Chao,Q., Chin,C., Chiou,J., Choi,E., Gonzalez,A., Howng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M., Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
JOURNAL	Unpublished		
REFERENCE	2	(bases 1 to 88643)	
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E., Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	3	(bases 1 to 88643)	
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E., Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P., Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
REFERENCE	4	(bases 1 to 88643)	
AUTHORS	Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct Submission		
JOURNAL	Submitted (12-SEP-2000) Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	On Feb 14, 2000 this sequence version replaced gi:6910565. Bases 57,901-88,643 of clone T28K15 overlap with bases 1-30,740 of IGF BAC clone F12F1 (AC002131). e-mail for correspondence: arab@sequence.stanford.edu		
	Genes with similarity to proteins in the databases are named 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The gene prediction programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev, http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark,http://www.cbs.dtu.dk/NetPlantGene.html).		
FEATURES	Location/Qualifiers		
source	1. .88643		
	/organism="Arabidopsis thaliana"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:3702"		
	/chromosome="I"		
	/map="near NCC1"		
	/clone="T28K15"		
misc_feature	1. .1849		
	/note="overlaps with BAC F5011 from 1-1849"		

gene	8039. .9467		
	/gene="T28K15.1"		
CDS	join(8039. .8198,8253. .8566,8712. .9467)		
	/gene="T28K15.1"		
	/note="Hypothetical protein"		
	/codon_start=1		
	/protein_id="AAG12568.1"		
	/db_xref="GI:10086508"		
	/translation="MNSFSHVPPGFRFHPPTDEELVDYYLRKKVASKRIEIDFIKIDIL YKIEPWDLQKILKVPFFYSVEIELCKICHEEQSDWYFFSHKDKKYPTGTRTNPATKA GFWKATGRDKAIYLRHSLIGMRKTLVFKGRAPNGQKSDWIMHEYRLETDENGTPQEE GWVVCRVFKRLAAVRMGDYDSSPSHWYDDQLSFMASELETNGQRRILPNHHQQQOH EHQQHMPYGLNASAYALNNPNLQCKQOELELHYNHLVQRNHLLDESHLSFLQLPLESP KIQQDNSNCNSLPYGTSDNIDNSSHNANLQQSNIAHEEQLNQGNQNFSSLYMNSGNEQ VMDQVTDWRVLDKFVASQLSNEEAATASASIQNNAKDTNAEYQVDEEKPDKRASDMG EEYTASTSSSCQIDLWK"		
	complement(11256. .13208)		
gene	/gene="T28K15.2"		
CDS	complement(join(11256. .11663,11757. .11791,12355. .12461, 12642. .12671,12775. .12942,13075. .13208))		
	/gene="T28K15.2"		
	/codon_start=1		
	/product="Hypothetical protein"		
	/protein_id="AAG12570.1"		
	/db_xref="GI:10086510"		
	/translation="WKSLDISRSSSVSRSPVHFQRYLLRRLQLSSRSNLEIKDSSNT SAESNTWKRILSAAWAAVIASSSGVPAMAELNRFADTRGEFGIGSAAQYGSADLSK TVHSNENFRANFTSADMRSDSFSGSTFNGAYLEKAVAYKANFGADLSDTLMDRMVL NEANLTNAVLRVSLTRSDLGAKIEGADFSDAVIDLLQKQVTTTHYIYPSFRSTIK KYFTNGFHNVLKALCKYATGTNPLTGVDTRKSLGCGNRRNAYGSPSSPLLSAPPQRL LGRDGFCDKGTGLCDVK"		
	complement(15068. .18826)		
gene	/gene="T28K15.3"		
CDS	complement(join(15068. .15286,15376. .15463,15549. .15961, 16068. .16685,17191. .17435,18087. .18095,18424. .18826))		
	/gene="T28K15.3"		
	/codon_start=1		
	/product="beta-fructosidase"		
	/protein_id="AAG12569.1"		
	/db_xref="GI:10086509"		
	/translation="MASSDALLPISAREEELCPYTRLPMDPNQETHGPRRRPFKG LLAVSFGLLFIAFYVALIATHDGSRSNDEGIDETETITSRARLAGVSEKRNGLWKLS GDRNTPAFENNSMLSWQRTAFHFQPEQNMNDPNGLFYKGWYHFFYQYNPNAAVWG DIVWGHAVSRDLIHWVLP IAMVADQWYDSNGVWTGSATFLPDGSIVMLYTGSTDKAV QVQNLAYPEDPNPLLKWKVPFGNVLVPPPLPDPFRDPTTAWKTSEGKWRITIG SKLNKTGISLVYDTIDFKTYEKLDTLLHRVPNTGWECVDFYPVSKTAGNGLDTSVNG PDVKHIVKASMDTRFDHYAVGTYFDSNGTWIPDDPTIDVGMTASLRDYDGKFASKS FYDQNKGRRLVWSWIGESDSEASDVQKGWSSLQGIPTVLDTKTGKNLVQWPVEIHK SLRLSSKQFDLEVGPGSVVPDVGSAQLDIEAFEINKESLDKIIGNASVVAEAEF SCEKSGGSTVRGALGPFSPVLATESLSEQTPVYFVAKGKDSCLKTFFCTDTSRSSV ANDVVKPIYGSVVPVLKGEKLTMRILVDHSIVAFPGQGGRTCITSRVYPTTAIYGAAK LFLFNALDATVTASTFTWQMNSAFIHPYSDEAVRALSRT"		
	complement(21817. .24475)		
gene	/gene="T28K15.4"		
CDS	complement(join(21817. .21963,22048. .22186,22271. .22446, 22520. .22589,22684. .22748,22824. .22904,22994. .23098, 23192. .23275,23370. .23445,23708. .23786,24190. .24331, 24422. .24475))		
	/gene="T28K15.4"		
	/note="Similar to transaldolase"		
	/codon_start=1		
	/protein_id="AAG12571.1"		
	/db_xref="GI:10086511"		
	/translation="MPLSLQSPPCATLSASIRKGRWKTAAAGFSVTPLPAVNFSLRRS IPRILASASSSSSPASSSLEAGENNELNAVSAFSEIVPDTVVFDFFERFPPTAATVSS ALLLGICGLPDTIFRNAVDMALADSSCAGLETTERSLSCFFNKAI VNVVGGDLVKLPVG RVSTEVDARLAYDTNGIIRKVHDLRLRYNEIDVPHDRLLFKIPATWQGI EAARLIESE GIQTHMTFVYSFAQAAASQAGASVIQIFVGRLDRWNRHSGDTEBSAISKSDPGL ALVKRSYNIHKYGYGSKLMAAAVRNKQDLFSLGLGVYVIAPLKLQSLKSPAI PDD EKYSFVRKLTPTETATHYHFTNKELVKWDQLSLASSMGPPASVELLSAGVEGYANQAKRV EELFGKIWPPPNV"		
	complement(24845. .27514)		
	/gene="T28K15.5"		



CDS	complement (24845.. .27514)	
	/gene="T28K15.5"	
	/note="NBS/LRR disease resistance gene, location of genomic clone gb AF074916."	
	/codon_start=1	
	/product="resistance to Pseudomonas syringae protein 5"	
	/protein_id="AAG12572.1"	
	/db_xref="GI:10086512"	
	/translation="MGGCFSVSLPCDQVVSQFSQLLCVRGSIYIHNLSKNLASLQKAMR MLKARQYDVIRRLTEEFTRGRQORLSQVQWLTSLVLIQNFNDLLRSNEVELQRLCL CGFCSKDLKLSYRYGKRVIMMLKEVESLSQGFVDVSEATPFADVDEIPFPQTTIVGQ EIMLEKAWNRLMEDSGILGLYGMGGVGKTTLLTKNNKFSKIDDRFDVVIWVVVSRs STVRKIORDIAEKVGLGMEWSEKNDNQIAVDIHNVLRRKFVLLDDDIWEKVNLIKAV GVPYPSKDNCGKVAFTTRSRDVCGRMGVDDPMEVSLQPEESWDLFQMKVKNTLGSsH PDIPGLARKVARKCRGLPALNVIGEAMACKRTVHEWCHAIIDLVTSSAIDPFSGMEDEI LHVLYSYDNLNGELMKSCFLYCSLFPEDYLLDKGLVDYWISEGFINEKEGRERNIN QGYEIIIGTLVRACILLLEERNKSNVKMHDVVREMAIWISSDLGKQKEKCIVRAGVGLR EVPKDKWNNTVRKISLMNNEIEEIFDSHECAALTTLFQKNDVVKISAEFFRCRMPHLV VLDLSENQSLNELPEEISELASLRYFNLSYTCIHQLPVGLWTLKKLIHLNLEHMSSLG SILGISNLWNLRITLGLDSRLLLDMSLVKELQLEHLEVITLDISSSLVAEPILLCSQR LVECIKEVDFKYLKEESVRVLTPTMGNLRKLGLKRCGMREIKIERTTSSSRNKSPsT TPCFNLRSRVFIAKCHGLKDLTWLLFAPNLTLFLBVGFsKEVEDIISEEKABEHSATIV PFRKLETLHLFELRGLKRIYAKALHPFCLKVIHVEKCEKLRKLPLDSKSGIAGEELVI YYGEREWIERVEWEDQTLRFLPSSRRWRRET"	
gene	complement (28920.. .31577)	
	/gene="T28K15.6"	
CDS	complement (28920.. .31577)	
	/gene="T28K15.6"	
	/note="Similar to RPS5, location of genomic clone gb AF074916."	
	/codon_start=1	
	/product="NBS/LRR disease resistance protein"	
	/protein_id="AAG12573.1"	
	/db_xref="GI:10086513"	
	/translation="MGGCVSVSLCDREVNQFOWMLCVSGSYIONLSENLASLQKAMG VLNAKRDDVQGRINREHEFTGHRRLAQVQVWLTRIQTIENQFNLLSTCNAEIQRLLC CGFCSKNVMSYLYGKRVIVLLREVEGLSQGFVDIVTEAAPIAEVEBELPIQSTIVGQ DSMLDKVWNCIMEDKVMIVGLYGMGGVGKTTLLTIQNNKFSKLGGGFVVIVVVVSKN ATVHKIQSIGEKLGLVGNWDEKNKNORALDIHNVLRRKKFVLLDDDIWEKVELKVI GVPYPSGENGCKVAFTTHSKEVCGRMGVDPNMEISCLDTGNAWDLLKKKGENTLGSsH PDIPLARKVSEKCCGLPALNVIGETMSFKRTIQEWRHATEVLTsATDFSMEDEIL PILKYSYDSLNGEDAKSCFLYCSLFPEDFEIRKEMLIYEWICEGFIKEKQGREKAFNQ GYDILGTLVRSSLLLEGAKDKDVVSMHDMVREMAIWFSDLGKHKERCIVQAGIGLDE LPEVENWRAVKRMSLMNNNPEKILGSPCEVELITLFLQNNYKLVDISMEFFRCMPSLA VLDLSENHSLSELPEEISELVSQYLDLSGTYIERLPHGLHELRLKLVHLKLEFRRLLE SIGISYLSLRTLRLRDSKTTLDTGLMKELQLLEHLELITTDISSGLVGELFCYPRV GRICIQHIIYIRDHWERPEESVGLVLPAPHNLCYTSIWNCWMMWEIMIEKTPMKKNLTNP NPSNLSNVRIEGCDGLKDLTWLLFAPNLINLRVWGCKHLEDIIISKEKAASVLEKEILP FQKLECLNLYQLSELKSIYWNALPFQRLRLCDILNCPKLRKLPLDSKSVVKVEEFVI KYKEKKWIERVEWEDEATQYRFLPTCRLR"	

Search completed: October 21, 2003, 09:29:30  
Job time : 18167 secs

Query Match		8.7%;	Score 448;	DB 8;	Length 88643;
Best Local Similarity		80.8%;	Pred. No. 7e-61;		
Matches 559;		Conservative	0;	Mismatches 130;	Indels 3; Gaps 3;
QY	1	TTTTTTGGCGGAAAAATCTC-GGGTTTACGTTTTTTGGCGGAAAAATCTCGTGTTTTACGTTT	59		
Db	78648	TTTTTTGGCGGAAAAATCACAGGATTTACATTTTGGCGGAAAAATCTTGGGTTTACGTTT	78589		
QY	60	TTGGCGGAAAAATCTCGGGTTTACGTTTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTT	119		
Db	78588	TTGGCGAGAAAAATCTTGGGTTTACGTTTTTTTTCGAGAAAAATCACGGGTTTAC-TTTTTTTG	78530		
QY	120	GCGGAAAAATCACGGATTATGTTTTTTTGGTGGAAAAAATTACGAGTTTACTTTTCTCAA	179		
Db	78529	GCGGAAAAATCACGGGTTTATGTTTTTTTGGTGGAAAAAATTACGAGTTTACTTTTCTCAA	78470		
QY	180	TTTCATTGCTTGATATTAAAGAAATTTGGAAAAATATTAAATTTTATAAATTGGTTTTAG	239		
Db	78469	TTTCATCGATTGTATATTAAAGAAATTTGGAAAAATATTAAATTTTATAAATTGGTTTTAG	78410		
QY	240	ATGTGTTGGTTAAACCTAAATTTGGCATTGGTTTAGAGATTTTAGTGGTTTATTCAATT	299		

Db	78409	ATGTGTTGGTTAAACTTAAATTGACAT	TGGTTTAGAGATTTTAGTTGGTTT	TATTCAAAT	78350
Qy	300	TTACAAAAATTTGATGGTTAATTGGAT	ATAAACCCATCGAAACCATTAACCA	CTTACAACTTAA	359
Db	78349	TTACAAAACTTATTGGGTTAATTGGGT	ATAAACCCATTAACCACTTACAACT	CAACCCAA	78290
Qy	360	CTCATTTTACTCATCAAAACCAATTGAC	TATCACTCACTCATCACTCACTTGTGA	419	
Db	78289	CTCATTTTACTCATCAAAACCAATTGAC	TATCACTCACTCATCACTCACTTGTGA	78230	
Qy	420	GTCAAAAAATTTCAACTCATTAGGGTT	CATGGTTGAGTTGAGTTGACCCATGA	479	
Db	78229	GTCAAAAAATTTTAACTCATTAGGGTT	CATGGATTGAGTTGAGTTGAGTTGA	78171	
Qy	480	TTTGACCCATTTTGACACCCCTACATAT	GATCATCAAGTTAAATAATCAAAAA	TACTATTG	539
Db	78170	TTTGACCCATTTTGACACCCCTAGTTT	AGAGTTATAGAAAACTGTAAAAAC	ACTAAAAATA	78111
Qy	540	ATAACTTTTAAACGGAAATTGTTTGT	TAAGTTTCATTTACGTTATTATATA	AAAAAATCAT	599
Db	78110	ATACATTTATAATCAACATTTAATAT	GTTTTCTTTAATATGTGTGTTTT	TCTAAACATCA	78051
Qy	600	AATGCAAAAGTACTAATGTATAGTT	TATTTTATTTTAAATAATGCAAAA	ATATTACTGT	659
Db	78050	AAACAAAAATGAACGGAGGAAGTACT	AAATTAAGTAATTAGTGAGAAAC	CTGTATTACCG	77991
Qy	660	ATACTTTTTCATTTCTTATCAATAT	TTTTTCTT	691	
Db	77990	CACTAGTTAAGCTCATAAAACTAAT	TCTCATT	77959	

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 00:52:07 ; Search time 9739 Seconds  
(without alignments)  
12837.275 Million cell updates/sec

Title: US-09-824-735-1  
Perfect score: 5144  
Sequence: 1 tttttggcgggaaaatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :		EST:*	
1:	em_estba:*		
2:	em_esthum:*		
3:	em_estin:*		
4:	em_estmu:*		
5:	em_estov:*		
6:	em_estpl:*		
7:	em_estro:*		
8:	em_htc:*		
9:	gb_est1:*		
10:	gb_est2:*		
11:	gb_htc:*		
12:	gb_est3:*		
13:	gb_est4:*		
14:	gb_est5:*		
15:	em_estfun:*		
16:	em_estom:*		
17:	em_gss_hum:*		
18:	em_gss_inv:*		
19:	em_gss_pln:*		
20:	em_gss_vrt:*		
21:	em_gss_fun:*		
22:	em_gss_mam:*		
23:	em_gss_mus:*		
24:	em_gss_pro:*		
25:	em_gss_rod:*		
26:	em_gss_phg:*		
27:	em_gss_vrl:*		
28:	gb_gss1:*		
29:	gb_gss2:*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	446.2	8.7	474	29	CNS00R7B
C 2	424.8	8.3	690	28	AQ961396
C 3	423	8.2	506	29	AL941773
C 4	375.2	7.3	433	29	CNS00SPA
					AL086741 Arabidops
					AQ961396 LERFL01TF
					AL941773 Arabidops
					AL088684 Arabidops

5	346.2	6.7	751	28	AQ961397	LERFL01TR
6	328	6.4	457	28	BH254517	SALK_0166
7	320.6	6.2	531	28	AQ966842	LERIM53TR
8	292.4	5.7	363	29	CC055323	SALK_0937
9	266	5.2	441	29	CNS00XC3	AL094689 Arabidops
10	259.8	5.1	319	28	BH902224	SALK_0914
C 11	231.4	4.5	528	28	BH237177	AUIIA63TR
12	216.6	4.2	367	29	BX289638	Arabidops
C 13	194.2	3.8	598	28	B77189	E77189 T3117TR TAM
14	185.8	3.6	514	9	AV830052	AV830052
C 15	173.6	3.4	454	29	BX291485	Arabidops
C 16	172.6	3.4	204	28	BH910603	SALK_0605
17	168.8	3.3	819	29	BZ456258	BOAO88TR
C 18	150.2	2.9	476	29	CC460177	CC460177 SALK_1392
19	150	2.9	816	28	BH599078	BH599078 BOGJY24TR
20	147.2	2.9	302	28	BH757375	BH757375 SALK_0561
C 21	147	2.9	672	28	B20959	B20959 T2H7-SP6 TA
C 22	121.8	2.4	133	29	BZ664600	SALK_0764
C 23	121.8	2.4	133	29	CC459381	SALK_1279
C 24	121.2	2.4	419	29	CC460016	CC460016 SALK_1384
25	115.2	2.2	168	28	BH905489	BH905489 SALK_1068
26	110.8	2.2	351	29	AL771439	AL771439 Arabidops
27	109.8	2.1	435	29	AL953182	AL953182 Arabidops
28	109.4	2.1	698	28	BH959869	BH959869 odi08h03
C 29	109	2.1	422	9	AV809816	AV809816
30	108.8	2.1	509	29	CNS000VAV	AL092053 Arabidops
31	108.8	2.1	589	10	BF005416	BF005416 EST433914
32	107.2	2.1	447	28	BH213915	SALK_0099
33	106.6	2.1	726	12	BI309472	BI309472 EST530882
34	105.6	2.1	584	10	BE801990	BE801990 sr23f02.y
35	105.2	2.0	740	12	BM985643	BM985643 2_F08_T3
36	104.4	2.0	670	13	BQ587842	BQ587842 S014307-0
C 37	103.6	2.0	1200	13	BX437758	BX437758 BX437758
C 38	103.2	2.0	404	14	CA798037	CA798037 CaC_BL_53
39	101.6	2.0	435	29	CC055495	CC055495 SALK_0949
40	101.4	2.0	506	13	BQ104557	BQ104557 fc1088.e
41	99.8	1.9	422	9	AI776493	AI776493 EST257593
42	99.8	1.9	492	9	AW738389	AW738389 EST339816
43	99	1.9	1201	9	AL514129	AL514129
44	98	1.9	401	12	BQ060265	BQ060265 Ts1316 Th
45	97.8	1.9	255	10	BE248163	BE248163 NF002C04D

ALIGNMENTS

RESULT 1	CNS00R7B/c	CNS00R7B	474 bp	DNA	linear	GSS 28-JUN-1999
LOCUS	Arabidopsis thaliana genome survey sequence SP6 end of BAC F13C22	Arabidopsis thaliana genome survey sequence SP6 end of BAC F13C22				
DEFINITION	of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.	of IGF library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.				
ACCESSION	AL086741	AL086741.1	GI:5287881			
VERSION	AL086741.1	GI:5287881				
KEYWORDS	GSS.	GSS.				
SOURCE	Arabidopsis thaliana (thale cress)	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana	Arabidopsis thaliana				
REFERENCE	1 (bases 1 to 474)	1 (bases 1 to 474)				
AUTHORS	Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.	Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.				
JOURNAL	Unpublished	Unpublished				
REFERENCE	2 (bases 1 to 474)	2 (bases 1 to 474)				
AUTHORS	Genoscope.	Genoscope.				
TITLE	Direct Submission	Direct Submission				
JOURNAL	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)	Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr)				
FEATURES	- Web : www.genoscope.cns.fr	- Web : www.genoscope.cns.fr				
source	Location/Qualifiers	Location/Qualifiers				
	1. .474	1. .474				
	/organism="Arabidopsis thaliana"	/organism="Arabidopsis thaliana"				

```

/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F13C22"
/clone_lib="IGF"
/note="end : SP6"

BASE COUNT      163 a      93 c      73 g      145 t
ORIGIN

Query Match      8.7%; Score 446.2; DB 29; Length 474;
Best Local Similarity 98.9%; Pred. No. 9.5e-46;
Matches 470; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 3274 AGTTTTTAATACATAGGATACTTGACCCCAATCCCAAAACAGTGAAGTATTTTGGCTTTGTT 3333
Db 474 AGTTTTTAATACATAGGATACTTGACCCCAATCCCAAAACAGTGAAGTATTTTGGCTTTGTT 415

QY 3334 CTCTCCTAGCTATCAGGTTTTTGGTGATATTTAATGTT-CTAGTAATATATATCTGTTTATC 3392
Db 414 CTCTCCTAGCTATCAGGTTTTTGGTGATATTTAAAGTTCTCTGTAATAATATCTGTTTATC 355

QY 3393 TATTATTGTTTCTCAATTAGAGCAGCGTANTCAAAATTCAGGAATCAAGAAATCAAGAAAGATCCTTG 3452
Db 354 TATTATTGTTTCTCAATTAGAGCAGCGTANTCAAAATTCAGGAATCAAGAAATCAAGAAAGATCCTTG 295

QY 3453 GTTCAGATTAAATTATGTGCCTATACGAGCAAGGGAAGAAAGAAAGTCAATTTGGATGA 3512
Db 294 GTTCAGATTAAATTATGTGCCTATACGAGCAAGGGAAGAAAGAAAGTCAATTTGGATGA 235

QY 3513 TATTCTGTCAGTTTTTGTATGGAATTGAGGTTTGTGCTTCGCCCTTCATATATGCTCTTTG 3572
Db 234 TATTCTGTCAGTTTTTGTATGGAATTGAGGTTTGTG- TTCGCCCTTCATATATGCTCTTTG 176

QY 3573 CTGGTCAATTCATTTAAATGTTAAGATCTCTTAGGAACGTTTGGATGACCAAGAAGAAG 3632
Db 175 CTGGTCAATTCATTTAAATGTTAAGATCTCTTAGGAACGTTTGGATGACCAAGAAGAAG 116

QY 3633 TGTGTTGCTACAGGATAGAACAAAATAGTAGCATGTGTTAGTTACCAAACCTGTAAACTG 3692
Db 115 TGTGTTGCTACAGGATAGAACAAAATAGTAGCATGTGTTAGTTACCAAACCTGTAAACTG 56

QY 3693 CTTCCTTTATTCAATTGCGCAACCATAGACCTTAGGAAGACTTAGAATTACAAGA 3747
Db 55 CTTCCTTTATTCAATTGCGCAACCATAGACCTTAGGAAGACTTAGAATTACAAGA 1

RESULT 2
AQ961396/c
LOCUS      AQ961396      690 bp      DNA      linear      GSS 28-JAN-2000
DEFINITION LERFL01TF LERA Arabidopsis thaliana genomic clone LERFL01, genomic
survey sequence.
ACCESSION  AQ961396
VERSION    AQ961396.1 GI:6789097
KEYWORDS  GSS.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 690)
AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms
JOURNAL   Unpublished
COMMENT   Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For additional information, see http://www.tigr.org/tdb/at/at.html


```

```

Seq primer: TF
Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..690
     /organism="Arabidopsis thaliana"
     /mol_type="genomic DNA"
     /strain="Landsberg erecta"
     /db_xref="taxon:3702"
     /clone="LERFL01"
     /clone_lib="LERA"
     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
sheared to 0.9-1 Kbp before ligation."

BASE COUNT      257 a      112 c      88 g      233 t
ORIGIN

Query Match      8.3%; Score 424.8; DB 28; Length 690;
Best Local Similarity 88.2%; Pred. No. 3.7e-43;
Matches 485; Conservative 0; Mismatches 62; Indels 3; Gaps 2;

QY 2 TTTTGGCGGAAAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGTGTTTACGTTTTT 61
Db 678 TTTTGGCGGAAAAATTACGGGTTTACGTTTTTGGCGGGAATTACGGGTTTACGTTTTT 619

QY 62 GGCGGAAAAATCTCGGTTTACGTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTC 121
Db 618 GGCGGAAAAATTATGAATTTACGTTTTTGGCGGAAGAAATCACGGGTTTAC-GTTTTTGGC 560

QY 122 GGGAAAAATCACGGATTATGTTTTTTTGGTGGAAAAATTACGAGTTTACTTTTTTCTCAATT 181
Db 559 GGGATAATCACGGGTTTATGTTTTTTGGTGGAAAAATTACGAGTTTATTTTTTCTCAATT 500

QY 182 TCATTGCTTGTATATTTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTAGAT 241
Db 499 TCATCGCTTGATATTTAAGAAATTTGGAAAAATATTAATTTTAAATTTGGTTAGAT 440

QY 242 GTGTTGTTAAACCTAAATTTGGCATTTGTTTAGAGATTTTAGTTGGTTTTTATTCAATTTT 301
Db 439 GTGTTGTTAAACTTAAATTTGGCATTTGTTTAAAGATTTTAGTTGGTTTAAATTCAATTTT 380

QY 302 ACAAAATTTGATGGTTAATTTGGATAAAACCATGGAAACCATTAACCATTACAACCTAACT 361
Db 379 ACAAAATTTGATGGTTAATTTGGTFAAACCATAGAAACCATTAACCATTACAACCCAAC 320

QY 362 CATTTTACTCATCAACCAATTTGACTCATCAACTCATTTGACTCATCAACTCATTTGAGT 421
Db 319 CA--TTACTCATCAACCAATTTGACTCATCAACCCATTTGACCCATCAACTCATTTGAGT 262

QY 422 CAAAAATTTCAAACCTCATTTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 481
Db 261 CAAAAATTTCAACTCATTTAGGTTTCATAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 202

QY 482 TGACCCATTTTGACACCCCTACATATGATCATAGTTAATTAATCAAAAAATTACTATTGAT 541
Db 201 TGACCCATTTTGACACCCCTAATATACTTTATTAAAAAGTAATAATTTGTAATCTTTAA 142

QY 542 AACTTTTTTAA 551
Db 141 AGCTAAATAA 132

RESULT 3
AL941773
LOCUS      AL941773      506 bp      DNA      linear      GSS 24-OCT-2002
DEFINITION Arabidopsis thaliana T-DNA flanking sequence GK-260F08-014949,
genomic survey sequence.
ACCESSION  AL941773
VERSION    AL941773.1 GI:24398371
KEYWORDS  GSS.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.


```

REFERENCE  
AUTHORS Strizhov,N., Li,Y., Rosso,M., Viehovever,P., Dekker,K., Saedler,H. and Weisshaar,B.  
TITLE A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.  
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics  
JOURNAL Unpublished  
REFERENCE  
AUTHORS Strizhov,N., Li,Y., Rosso,M. and Weisshaar,B.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-2002) Weisshaar B., Max-Planck-Institut fuer Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
COMMENT This sequence is recovered from the left border of the T-DNA. It indicates an insertion close to or within gene At5g35410. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:  
http://www.mpiz-koeln.mpg.de/GABI-Kat/.  
FEATURES  
source Location/Qualifiers  
1. .506  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="GK-260F08-014949"  
/clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"  
BASE COUNT 140 a 86 c 103 g 177 t  
ORIGIN  
Query Match 8.2%; Score 423; DB 29; Length 506;  
Best Local Similarity 96.3%; Pred. NO. 6.8e-43;  
Matches 488; Conservative 0; Mismatches 10; Indels 9; Gaps 5;  
QY 3939 TCTTCTAAACCTTGACTTCAATTGTTAAACTAAACCTCAGTATCTGTCTTAGCTAAAG 3998  
Db 1 TCTTCAAAAACCTTGACTTAATTGTT---AACTAAACCTCAGTATCTGTCTTAGCTAAAG 57  
QY 3999 TTACTTTTACTTGTGTTTTTCATTAAGTTGACCTGTCAATTGCACTTGTTTACAGGGCAGTT 4058  
Db 58 -TACTTTTACTTGTTTTA---TTAGTTGACCTGT-AAATTGCACTTGTT-ACAGGGCAGTT 111  
QY 4059 ATGTAGCGGAGAAATGTAGAGAGAGAAATGATGAAGGCCCCCTGATGATGAATGCCTTTGAGA 4118  
Db 112 ATGTAGCGGAGAAATGTAGAGAGAAATGATGAAGGCCCCCTGATGATGAATGCCTTTGAGA 171  
QY 4119 TGATTACCTTATCACAAGGCTTAAATTTATCTGCACTATTTTGACAGGGCGACAGGTAGTAC 4178  
Db 172 TGATTACCTTATCACAGGCTTAAATTTATCTGCACTATTTTGACAGGGCGACAGGTAGTAC 231  
QY 4179 CTGATTTTCTATTACTGGTCATAGAGATCTCCATTTCCGAATAAAAGAATGTCGGTAGCAT 4238  
Db 232 CTGATTTTCTATTACTGGTCATAGAGATCTCCATTTCCGAATAAAAGAATGTCGGTAGCAT 291  
QY 4239 CTATTTCTTCAGACTGCCGTTTTGACTGCCCTTATGATGCTGTGTTCTTAGTTGTATATAA 4298  
Db 292 CTATTTCTTCAGACTGCCGTTTTGACTGCCCTTATGATGCTGTGTTCTTAGTTGTATATAA 351  
QY 4299 TAACTATAAGTTCATTAGATGATTGGTTGCAITGGCATTAGTAGATACAAATGGAATCCAA 4358

Db 352 TAACTATAAGTTCATTAGATGATTGGTTGCATGGCATTAGTAGATACAAATGGAATCCAA 411  
QY 4359 AATGTTCTCTGCATATTGATGGCTGATCCTTTTGATCTCGCAGGATTTGTTAAAAAGGCAAA 4418  
Db 412 AATGTTCTCTGCATATTGATGGCTGATCCTTTTGATCTCGCAGGATTTGTTAAAAAGGCAAA 471  
QY 4419 CCCGTTTTGTTTCTCGAAGGGGAACCTA 4445  
Db 472 CCCGTTTTGTTTCTCGAAGGGGAACCTA 498  
RESULT 4  
CNS00SPA/c 433 bp DNA linear GSS 28-JUN-1999  
LOCUS Arabidopsis thaliana genome survey sequence SP6 end of BAC T2H7 of  
DEFINITION TAMU library from strain Columbia of Arabidopsis thaliana, genomic survey sequence.  
ACCESSION AL088684  
VERSION AL088684.1 GI:5289824  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 433)  
AUTHORS Salanoubat,M., Choisine,N., Artiguenave,F., Brottier,P., Wincker,P., Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 433)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
FEATURES  
source Location/Qualifiers  
1. .433  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T2H7"  
/clone\_lib="TAMU"  
/note="end : SP6"  
BASE COUNT 162 a 75 c 66 g 130 t  
ORIGIN  
Query Match 7.3%; Score 375.2; DB 29; Length 433;  
Best Local Similarity 93.3%; Pred. No. 5.6e-37;  
Matches 403; Conservative 0; Mismatches 28; Indels 1; Gaps 1;  
QY 79 TTTACGTTTTTTCGAGAAAAATCACGGGTTTACTTTTTTTTTCGGGAAAAATCACGGATTT 138  
Db 433 TTTACGTTTTTTCGGGGAAAAATCACGGGTTTAC-GTTTTTGGCGGAAAAATCACGGGTTT 375  
QY 139 ATGTTTTTTTGGTGAAAAAATTACGAGTTTACTTTTTCTCAATTTCAATGCTTGTATATTT 198  
Db 374 ATGTTTTTTTGGTAGAAAAATTACGAGTTTACTTTTTCTCAATTTCAATGCTTGTATATTT 315  
QY 199 AAGAAATTTGGAAAAATATTAATTTTATAAATTGGTTTAGATGTTGGTTAAACCTAA 258  
Db 314 AAGAAATTTGGAAAAATATTTGATTTTATAAATTGGTTTAGATGTTGGTTAAGCTTAA 255  
QY 259 ATTGGCATTGGTTTAGAGATTTTAGTGGTTTTTATTCAATTTTACAAAAATTTGATGGGTT 318  
Db 254 ATTGGCATTGGTTTAGAGATTTTACTTGGTTTAATTTACAAAACTTGTATGGGTT 195  
QY 319 AATTGGATAAACCATGGAAACCATTAACCATTAACAACCTAATCTATCTTACTCATCAAAC 378  
Db 194 AATTGGGTAACCATTTGAAACCATTAACCATTAACAACCTAATCTATCTTACTCATCAAAC 135  
QY 379 CAATTGACTCATCAACTCATTTGACTCATCAACTCATTTGAGTCAAAAAATTTCAACTCAT 438





QY 4699 AATTGTTATGATTTGTTACGTTCTATGTGCTCAACAGACAAGGCTCGAGGGATTATCTTCG 4758  
|||||  
Db 141 AATTGATATGATTTGTTACGTTCTATGTGCTCAACAGACAAGGCTCGATGGATTATCTTCA 200  
QY 4759 ATCAAGGCCGGACAGTTAGCTGTTGTGATAGAGGTAAATTATGCTTGTGTTGATGTAAT 4818  
|||||  
Db 201 ATCAAGGCCGGACAGTTAGCTGTTGTGATAGAGGTAAATTATGCTTGTGTTGATGTAAT 260  
QY 4819 ATAAGTTTGCTTTGCTTCAGTTTAAGGGTATCTAGCAAAATTGAAATTAACCTACATGCAG 4878  
|||||  
Db 261 ATAAGTTTGCTTTGCTTCAGTTTAAGGGTATCTAGCAAAATTGAAATTAACCTACATGCAG 320  
QY 4879 ATTTACGAGGTGGCACCATCGCTTTTCATGGTAGACGTAAGAAAAGGCTGCTGGTGAAAAC 4938  
|||||  
Db 321 ATTTACGAGGTGGCAGCCCGGTCGTCGACCACGCGTGCCCTATGGTGAGTCGTAAAAACT 380  
QY 4939 CTTGAATATCACAAAGTTTAAATATATATATCCAAATAACAATAGTTGCAATCACTACTGTG 4998  
|||||  
Db 381 CTTGAATATCGCAAGGCTTATCA--TACTATCCAATGACAGTAGTTGCTTCATCACTGTC 438  
QY 4999 TTCCGGATTAGAGTGATA 5016  
|  
Db 439 GCTCGCGTTAGAGTGATA 456

RESULT 7  
AQ966842 531 bp DNA linear GSS 28-JAN-2000  
LOCUS LERIM53TRB LERG Arabidopsis thaliana genomic clone LERIM53, genomic  
survey sequence.  
ACCESSION AQ966842  
VERSION AQ966842.1 GI:6794543  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 531)  
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,  
Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.  
TITLE Genomic survey sequencing of Landsberg erecta ecotype of  
Arabidopsis thaliana and identification of sequence-based  
polymorphisms  
JOURNAL Unpublished  
COMMENT Contact: Xiaoying Lin  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: at@tigr.org  
For additional information, see <http://www.tigr.org/tdb/at/at.html>  
Seq primer: TR  
Class: shotgun.

FEATURES source  
1..531  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Landsberg erecta"  
/db\_xref="taxon:3702"  
/clone="LERIM53"  
/clone\_lib="LERG"  
/note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was  
sheared to 0.4-0.7 Kbp before ligation."  
BASE COUNT 147 a 68 c 117 g 199 t

ORIGIN  
Query Match 6.2%; Score 320.6; DB 28; Length 531;  
Best Local Similarity 91.0%; Pred. No. 2.9e-30;  
Matches 352; Conservative 0; Mismatches 34; Indels 1; Gaps 1;

QY 1 TTTTGGCGGAAAAATCTCGGGTTTACGTTTGGCGGAAAAATCTCGTGTACGTTTT 60  
|||||

Db 146 TTTTGGCAGAAAAATCACGGGTTTACGTTTGGCAGAAAAATTATGAATTACGTTAT 205  
QY 61 TGGCGGGAAAAATCTCGGGTTTACGTTTGGCAGAAAAATCACGGGTTTACGTTTGG 120  
|||||  
Db 206 TGGCGGGAAAAATCTCGGGTTTACGTTTGGCGGAAAAATCACGGGTTTAC-GTTTTGG 264  
QY 121 CGGGAAAAATCACGGAATTATGTTTTTTGGTGGAATAATACGAGTTTACITTTTCTCAAT 180  
|||||  
Db 265 CGGGAAAAATCACGGGTTTATGTTTTTTTAAAGAAAAATTAATAATTACTAGTTTACTTTTCTCAAT 324  
QY 181 TTCATTGCTTGTATATTTAAGAAATTTGGAATAATTAATTTTATAAATCGTTTAGA 240  
|||||  
Db 325 TTCATCGCTTGTATATTTAAGAAATTTGGAATAATTAATTTTATAAATCGGCTTAGA 384  
QY 241 TGTGTTGGTTAAACCTAAATTTGGCATTGGTTTAGAGATTTTAGTTGGTTTATCAATTT 300  
|||||  
Db 385 TGTGTTGGCTAAACTAAATTTGGCATTGGTTTAGAGATTTTAGTTGGTTTAAATCAATTT 444  
QY 301 TACAAAAATTTGATGCGTTAATTTGGATAAACCATGGAACCATTAACCATTAACCTAAC 360  
|||||  
Db 445 TACAAAACTTGATGCGTTAATTTGGTAAACCATTTGGAACCATTAACCATTAACCCACC 504  
QY 361 TCATTTTACTCATCAACCAATTGACT 387  
|  
Db 505 TTATTTTACTCATCAACCAATTGACT 531

RESULT 8  
CC055323 363 bp DNA linear GSS 02-APR-2003  
LOCUS SALK\_093737.53.25.x Arabidopsis thaliana TDNA insertion lines  
DEFINITION Arabidopsis thaliana genomic clone SALK\_093737.53.25.x, genomic  
survey sequence.  
ACCESSION CC055323  
VERSION CC055323.1 GI:29474987  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids  
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 363)  
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab  
., C. Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,  
Zimmerman,J. and Ecker,J.R.  
TITLE A Sequence-Indexed Library of Insertion Mutations in the  
Arabidopsis Genome  
JOURNAL Unpublished  
COMMENT Contact: Joseph R. Ecker  
Salk Institute Genomic Analysis Laboratory (SIGnAL)  
The Salk Institute for Biological Studies  
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA  
Tel: 858 453 4100 x1752  
Fax: 858 558 6379  
Email: ecker@salk.edu  
This is single pass sequence recovered from the left border of  
TDNA.

FEATURES source  
Location/Qualifiers  
1..363  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="SALK\_093737.53.25.x"  
/clone\_lib="Arabidopsis thaliana TDNA insertion lines"  
/note="PCR was performed on Arabidopsis thaliana lines  
each of which contains one or more TDNA insertion  
elements. The resultant fragment for each line was  
directly sequenced to determine the genomic sequence at  
the site of insertion. Details of the protocols used can  
be found at [http://signal.salk.edu/tdna\\_protocols.html](http://signal.salk.edu/tdna_protocols.html)"  
BASE COUNT 115 a 58 c 57 g 132 t

Db	48	ACGAGTTTAAATCTCTCAATTTCAATTCATTCATATATATATTAAGAAATTCGAAAAATATTTG	107
QY	220	ATTTTATTAAATTTGGTTTAGATGTGTTGTTAAACCTAAATTTGGCATTTGGTTTAGAGATT	279
Db	108	ATTGTATTAAATTTGGTTTAGATGTGTTGTTAAACTTAAATCGGCATTTGGTTTAGAGATT	167
QY	280	TTAGTTGGTTTATTTCAATTTTACAAAATTTTGATGGGTTAAATTTGGATAAACCATGGAAAC	339
Db	168	TTAGTTGGTTTATTTATTTTACAAAACTTGATGGGTTAATTTGGGTA-----	215
QY	340	CATTAAACCATTTACACCTAACTCATTTTACTCATCAAAACCAATTTGACTCATCAACTCATT	399
Db	216	----AACCATTTACACCCAACTCATTTTATTTATTCATAACCAATTTGACTCAATACTCATT	271
QY	400	TGACTCATCAACTCATTTTGAGTCAAAAATTTTCAAACTCATAGGGTTTCATGGGTTGAGTTG	459
Db	272	TGACCCATCAACTCATTTTGAGTTAATATTTTCAAACTCATAGGGTTTCATGAGTTGAGTTG	331
QY	460	AGTTGAGTTGACCCATGAATTTTGACCCCATTTTGACACCCCTACATATGATCATAGTTA	519
Db	332	AGTTGAGTTGACCCATGAATTTTGATCCATTTTGACACCCCTAGGCCCGTCTCATATGAGA	391
QY	520	ATAATCAAAAATTTACTATTTGATAAACTTTTAAACGGAATTTGTTT	563
Db	392	TTTAAACGACCATGTCATGGTTAGCAACCACTGAATATTTT	435
RESULT 10			
BH902224			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
QY	143	TTTTTGGTGGAAAAATTACGAGTTTACTTTTCTCAATTTCAATTCATTCGTTGTATATTAAGA	202
Db	1	TTTTTGGTGGATAAAATTACGAGTTTACGCTTCTCAATTTCAATTCATCTCTTGATATTAAGA	60
QY	203	A-ATTTGGAAAAATATTAATTTTATTAATTTGGTTTAGATGTGTTGTTAAACCTTAAAT	261
Db	61	ANATTTGGAAAAATATTTGATTTTCTTAAATTTGGTTTAGATGTGTTTAAACITAAAT	120
QY	262	GGCATTGGTTTAGAGATTTTAGTTGGTTTATTTCAATTTTACAAAATTTGATGGGTTAAT	321
Db	121	GATATTGGTTTAGAGATTTTATTGGTTTAAATTTCAATTTTACAAAATTTGATGGGTTAAT	180
QY	322	TGGATAAACCATGGAAACCAATTAAACCATTAACCTAATTTTACTCATCAACCAAA	381
Db	181	TGGTAAACCATTTGAAACCAATTAAACCATTAACCCCACTCATTTTACTCATCAACCAAA	240
QY	382	TTGACTCATCAACTCATTTGACTCATCAACTCATTTGAGTCAAAAATTTCAACTCATTAG	441
Db	241	TTGACCCCAATAACTCATTTGACCCCATCAACTCATTTGAGTCAAAAATTTCAACTAATAG	300
QY	442	GGTTCATGGGTTGAGTTGAGTTGAGTTGACCCCATGAATTTTGACCC	487
Db	301	GAATCATGGGTTGAGTTGAGTTGAGTTGACCCCATGAATTCGAGCCC	346
RESULT 9			
CNS00XC3			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT			
ORIGIN			
QY	160	ACGAGTTTACTTTTCTCAATTTTCAATTCATTCGTTGTATATTTAAGAAATTTGAAAAATATTA	219

ORIGIN									
Query Match		5.1%;		Score 259.8;		DB 28;		Length 319;	
Best Local Similarity		95.7%;		Pred. No. 1.1e-22;					
Matches 267;		Conservative		0;		Mismatches 12;		Indels 0; Gaps 0;	
QY	229	AATGGTTTAGATGTGTGGTTAAACCTAAATTGGCATTGGTTTAGAGATTTAGTGGT 288							
Db	1	AATGGTTTAGATGTGTGGTTAAACTTAAATTGGCATTGGTTTAGAGATTTAGTGGT 60							
QY	289	TTTATTCAATTTTACAAAATTTGATGGGTTAATTGGATAAAACCATGGAACCATTAACCA 348							
Db	61	TTTATTCAATTTTACAAAATTTGATGGGTTAATTGGGTAAACCATTTGAAACCATTAACCA 120							
QY	349	TTACAACCTAACTCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACTCATC 408							
Db	121	ATACAACCAACTCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTTGACCCATC 180							
QY	409	AACTCATTTGAGTCAAAAATTTCAACTCATTTAGGGTTTCATGGGTTGAGTTGAGTTGAGTT 468							
Db	181	AACTCATTTGAGTCAAAAATTTTCAACTCATTTAGGGTTTCATGGGTTGAGTTGAGTTGAGTT 240							
QY	469	GACCCATGAATTTTGACCCATTTTGACACCCCTACATAT 507							
Db	241	GACTCATGAATTTTGACCCATTTTGACACCCCTAACAAAT 279							
RESULT 11									
BH237177/c									
LOCUS									
DEFINITION									
BH237177 528 bp DNA linear GSS 13-NOV-2001									
survey sequence.									
BH237177									
BH237177.1 GI:16907548									
GSS.									
Arabidopsis thaliana (thale cress)									
Arabidopsis thaliana									
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;									
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids									
; eurosids II; Brassicales; Brassicaceae; Arabidopsis.									
1 (bases 1 to 528)									
Town, C.D., Whitelaw, C.A., Pai, G., Van Aken, S.E., Utterback, T.V.,									
Feldblyum, T.V. and Fraser, C.M.									
Survey sequencing of Arabidopsis thaliana BAC T2P24									
Unpublished									
Other GSSs: AUIIA63TF									
Contact: Chris Town									
TIGR									
9712 Medical Center Drive, Rockville, MD 20850, USA.									
Tel: 301-838-3523									
Fax: 301-838-0208									
Email: cdtown@tigr.org									
From Wash. U contig 849.									
Seq primer: TR									
Class: sheared ends.									
Location/Qualifiers									
1..528									
/organism="Arabidopsis thaliana"									
/mol_type="genomic DNA"									
/strain="Columbia"									
/db_xref="taxon:3702"									
/clone="AUIIA63"									
/clone_lib="AUII"									
/note="Vector: pHOS2; Site 1: BstXI; 2-3 kb sheared BAC									
DNA inserted into pHOS2 using BstXI linkers"									
BASE COUNT		222 a		99 c		51 g		156 t	
ORIGIN									
Query Match		4.5%;		Score 231.4;		DB 28;		Length 528;	
Best Local Similarity		87.9%;		Pred. No. 2.9e-19;					
Matches 275;		Conservative		0;		Mismatches 36;		Indels 2; Gaps 2;	
QY	1	TTTTTGGCGGAAAAATCTCGGGTTTACGTTTACGTTTGGCGGAAAAATCTCGTTTACG-TTT 59							

Db	313	
----	-----	--





An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al., 1998). cDNA cleaved with BamHI and XhoI was ligated to modified lambda FLC-1 vector. (Carninci et al., submitted for publication) digested with BamHI and Sali. This clone is in a modified pBluescript vector. Please visit our web site ([http://www.gsc.riken.go.jp/e/plant/index\\_e.html](http://www.gsc.riken.go.jp/e/plant/index_e.html)) for further details.

FEATURES  
source Location/Qualifiers  
1..514  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/clone="RAFL09-61-G03"  
/dev\_stage="plants at various developmental stages from germination to mature seeds"  
/lab\_host="DH10B"  
/clone\_lib="RAFL9"  
/note="Site 1: BamHI; Site 2: Sali; subjected to dehydration (1, 2, 5, 10, 24 hr) and cold (1, 2, 5, 10, 24 hr) treatments"  
BASE COUNT 163 a 88 c 123 g 137 t 3 others  
ORIGIN

Query Match 3.6%; Score 185.8; DB 9; Length 514;  
Best Local Similarity 98.9%; Pred. No. 1.3e-13;  
Matches 187; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1983 TCAGATAAAAGTTTGTAAAGATGACAAAGAAAATGAGAGAGTGGGCAAGTACGAGGTT 2042  
Db 77 TCTGATAAAAGTTTGTAAAGATGACAAAGAAAATGAGAGAGTGGGCAAGTACGAGGTT 136  
QY 2043 GGTCGCACAATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTGGAGGAACACAGACACT 2102  
Db 137 GGTCGCACAATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTGGAGGAACACAGACACT 196  
QY 2103 GGTGATAATGTAGCCATCAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAATGGTT 2162  
Db 197 GGTGATAATGTAGCCATCAAATTTATGGCTAAGAGTACAATACTTAAGAACAGAATGGTT 256  
QY 2163 GATCAGGTA 2171  
Db 257 GATCAGATA 265

RESULT 15  
BX291485/c  
LOCUS  
DEFINITION BX291485 454 bp DNA linear GSS 07-MAR-2003  
Arabidopsis thaliana T-DNA flanking sequence GK-463G11-018953,  
genomic survey sequence.  
ACCESSION BX291485  
VERSION BX291485.1 GI:28890481  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi  
1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H.  
and Weisshaar,B.  
A pipeline for automated high-throughput generation of FSTs  
(flanking sequence tags) from Arabidopsis thaliana T-DNA  
transformed lines  
Unpublished  
2  
JOURNAL  
REFERENCE  
AUTHORS Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weisshaar,B.  
TITLE A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat)  
for flanking sequence tag based reverse genetics  
Unpublished  
3 (bases 1 to 454)  
JOURNAL  
REFERENCE  
AUTHORS Li,Y., Strizhov,N., Rosso,M. and Weisshaar,B.  
TITLE Direct Submission  
JOURNAL Submitted (07-MAR-2003) Weisshaar B., Max-Planck-Institut fuer

COMMENT

Zuechtungsforschung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany  
This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone F9C16. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at:  
<http://www.mpiz-koeln.mpg.de/GABI-Kat/>.

FEATURES  
source Location/Qualifiers

1..454  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/strain="Columbia 0"  
/db\_xref="taxon:3702"  
/clone="GK-463G11-018953"  
/clone\_lib="Arabidopsis thaliana T-DNA insertion lines"  
/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector pAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"

BASE COUNT 171 a 91 c 45 g 147 t  
ORIGIN  
Query Match 3.4%; Score 173.6; DB 29; Length 454;  
Best Local Similarity 77.2%; Pred. No. 4.2e-12;  
Matches 227; Conservative 0; Mismatches 59; Indels 8; Gaps 1;  
QY 179 ATTTCAATTGCTTGATATTTAAAGAAATTTGGAATAATATTAATTTTATTAAATTCGTTTA 238  
Db 306 AATAAATTGATGGGTTGGTGAAGCCATTGTGAAAAAGATTGATTATATTCAATTGATTTA 247  
QY 239 GATGTGTTGGTTAAACCTAAATTGGCATTGGTTTACAGATTTTAGTTGGTTTATTCAAT 298  
Db 246 GAGGTTTTGGTTAAACTCAAATTGATATCGATTTAGATTGGATTAAATTCAAT 187  
QY 299 TTTACAAAATTTGATGGTTAATTGGATAAACCAATGGAACCAATTAACCAATTAACAACCTA 358  
Db 186 TTAGCCAAAATTGGTGGGTTAAATGGGTAAACCAT-----TAACCCATTGAAACCCA 135  
QY 359 ACTCATTTTACTCATCAAAACCAATTGACTCATCAACTCATTCATCAACTCATTTG 418  
Db 134 ACTCATTGATCCATCTAACCAATTTGACCCCAATAAACTCATTTGAACTATCAATTCATTG 75  
QY 419 AGTCAAAAATTTCAACTCATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGACC 472  
Db 74 AGTGAAAAATTTTCATCTCATTAGGGTTTCATGGGTTGAGTTGAATTGGGTTGATC 21

Search completed: October 21, 2003, 04:26:39  
Job time : 9747 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 04:26:42 ; Search time 1236 Seconds  
(without alignments)  
11161.152 Million cell updates/sec

Title: US-09-824-735-1  
Perfect score: 5144  
Sequence: 1 tttttggcgggaaatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5144	100.0	5144	9	US-09-824-735-1
2	414.6	8.1	2000	10	US-09-938-842A-2784
3	401.4	7.8	2000	10	US-09-938-842A-5327
4	371.2	7.2	2000	10	US-09-938-842A-4657
5	152.4	3.0	2000	10	US-09-938-842A-5099
6	97.6	1.9	404	10	US-09-878-574-3933
7	95.6	1.9	1468	12	US-10-292-408-8
8	91.6	1.8	3673778	12	US-10-312-841-1
9	88	1.7	742	9	US-09-770-149-51
10	83.4	1.6	3673778	12	US-10-312-841-2
11	82.2	1.6	1326	10	US-09-938-842A-72
12	81.8	1.6	15548	12	US-10-311-455-2128
13	80.4	1.6	15387	12	US-10-311-455-157
14	77.4	1.5	3673778	12	US-10-312-841-2
15	76.6	1.5	7306	12	US-10-311-455-1610
16	75.2	1.5	1326	10	US-09-938-842A-2313

17	74.6	1.5	1470	10	US-09-938-842A-1978
18	74.2	1.4	5893	12	US-10-311-455-832
19	73.8	1.4	464	9	US-09-770-444-356
20	73.8	1.4	1290	10	US-09-938-842A-1831
21	73.8	1.4	6314	12	US-10-240-452-14
22	72.4	1.4	1266	10	US-09-938-842A-1288
23	72.4	1.4	18218	12	US-10-311-455-1922
24	72.2	1.4	6030	12	US-10-240-453-186
25	72.2	1.4	6030	14	US-10-239-676-164
c	71.4	1.4	461	12	US-09-814-353-17724
c	71.4	1.4	2000	10	US-09-938-842A-3403
28	70.8	1.4	7002	12	US-10-311-455-783
29	70.8	1.4	10480	12	US-10-311-455-2173
30	70.4	1.4	16724	12	US-10-311-455-1063
31	70.4	1.4	16724	12	US-10-240-485-89
32	70.2	1.4	393	10	US-09-960-352-4582
33	70	1.4	6419	12	US-10-311-455-240
34	70	1.4	17183	12	US-10-311-455-460
35	69.2	1.3	5689	12	US-10-240-453-100
36	69.2	1.3	5689	14	US-10-239-676-90
c	69.2	1.3	6590	12	US-10-311-455-1449
38	69.2	1.3	7143	12	US-10-311-455-956
39	69	1.3	480	10	US-09-960-352-5301
40	68.6	1.3	1281	10	US-09-938-842A-70
41	68.2	1.3	6341	12	US-10-311-455-1617
42	68.2	1.3	9707	12	US-10-311-455-1393
43	68	1.3	9539	12	US-10-240-453-53
44	68	1.3	9539	14	US-10-239-676-51
45	68	1.3	12705	12	US-10-311-455-122

ALIGNMENTS

RESULT 1  
US-09-824-735-1  
; Sequence 1, Application US/09824735  
; Patent No. US20020095032A1  
; GENERAL INFORMATION:  
; APPLICANT: ZHU, JIAN-KANG  
; APPLICANT: LIU, JIPING  
; APPLICANT: ISHITANI, MANABU  
; APPLICANT: HALFTER, URSULA  
; APPLICANT: KIM, CHEOL-SOO  
; TITLE OF INVENTION: PROTEINS AND DNA RELATED TO SALT TOLERANCE IN PLANTS  
; FILE REFERENCE: 205645US20  
; CURRENT APPLICATION NUMBER: US/09/824,735  
; CURRENT FILING DATE: 2001-08-17  
; PRIOR APPLICATION NUMBER: US 60/824,735  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 5144  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2004)..(2168)  
; OTHER INFORMATION:  
; NAME/KEY: CDS  
; LOCATION: (2255)..(2317)  
; OTHER INFORMATION:  
; NAME/KEY: CDS  
; LOCATION: (2416)..(2487)  
; OTHER INFORMATION:  
; NAME/KEY: CDS  
; LOCATION: (2573)..(2680)  
; OTHER INFORMATION:  
; NAME/KEY: CDS  
; LOCATION: (2778)..(2852)  
; OTHER INFORMATION:  
; NAME/KEY: CDS





Db 1741 ATTTATATATTGATTATGATCTTGATCCTTACCTTTTCGTCGTTTCCTTGATCGAATTT 1800  
QY 1801 ATGTTGATTTATGCTGCGAAATCAAAAATTATGGATACAGAGTAATTTTGTATATATGAT 1860  
Db 1801 ATGTTGATTTATGCTGCGAAATCAAAAATTATGGATACAGAGTAATTTTGTATATATGAT 1860  
QY 1861 GTGTAGGAGCTAAAAATTAGCAAAACAATACGTAAGTAATTGAAATCGAAATCATAAACGT 1920  
Db 1861 GTGTAGGAGCTAAAAATTAGCAAAACAATACGTAAGTAATTGAAATCGAAATCATAAACGT 1920  
QY 1921 TTAAGGAAAAGAGGTTTTTTACTAAAGTCTCTGAATAATCTGATGTATAGCTTGTGGTCAATG 1980  
Db 1921 TTAAGGAAAAGAGGTTTTTTACTAAAGTCTCTGAATAATCTGATGTATAGCTTGTGGTCAATG 1980  
QY 1981 GATCAGATAAAAGTTTTGTAAAGAAATGACAAAGAAAATGAGAAGAGTGGGCAAGTACGAGG 2040  
Db 1981 GATCAGATAAAAGTTTTGTAAAGAAATGACAAAGAAAATGAGAAGAGTGGGCAAGTACGAGG 2040  
QY 2041 TTGGTCGCACAATAGGTGAAGAACCTTTTGCTAAGGTTAAGTTTTGCGAGGAACACAGACA 2100  
Db 2041 TTGGTCGCACAATAGGTGAAGAACCTTTTGCTAAGGTTAAGTTTTGCGAGGAACACAGACA 2100  
QY 2101 CTGGTGATAAATGTAGCCATCAAAAATTATGGCTAAGAGTACAATACTTAAAGAACAGAAATGG 2160  
Db 2101 CTGGTGATAAATGTAGCCATCAAAAATTATGGCTAAGAGTACAATACTTAAAGAACAGAAATGG 2160  
QY 2161 TTGATCAGGTATGTTCTGGATGTGTTTTTACATGGAACTAAGGTTGTTGCGTCAATGGT 2220  
Db 2161 TTGATCAGGTATGTTCTGGATGTGTTTTTACATGGAACTAAGGTTGTTGCGTCAATGGT 2220  
QY 2221 ATGATCTTTGATTTTCGTTTTAAAGCTCTTTTACAGATAAAAAGAGAGATATCTATAATGAA 2280  
Db 2221 ATGATCTTTGATTTTCGTTTTAAAGCTCTTTTACAGATAAAAAGAGAGATATCTATAATGAA 2280  
QY 2281 GATTGTTCGTCACCCGAAACATAGTGAGGTTGTATGAGGTAATGTTTGTTCCTATGCA 2340  
Db 2281 GATTGTTCGTCACCCGAAACATAGTGAGGTTGTATGAGGTAATGTTTGTTCCTATGCA 2340  
QY 2341 TCTGCGAAATTTTATCTCTGAAGTGTTTTGTGCATCATTTGTTCTCTGTGTTTTTTTGTG 2400  
Db 2341 TCTGCGAAATTTTATCTCTGAAGTGTTTTGTGCATCATTTGTTCTCTGTGTTTTTTTGTG 2400  
QY 2401 ATTTTCCCGATGTAGGTTGTCGAGTCTTTTGCATCATTTGTTCTCTGTGTTTTTTTGTG 2460  
Db 2401 ATTTTCCCGATGTAGGTTGTCGAGTCTTTTCGAAAATATATATAGTTTTTGGAGTTTTGTG 2460  
QY 2461 ACAGGAGGAGAGCTCTTTGTATAGAATTTGTACGGAACTTCCATACTTGTAGGCAGCGTCCA 2520  
Db 2461 ACAGGAGGAGAGCTCTTTGTATAGAATTTGTACGGAACTTCCATACTTGTAGGCAGCGTCCA 2520  
QY 2521 TTAGTTAAAAACCTCTCTACTTAATTTTTTAAATATATAGAAATCTTTTCATGCAGGTTTCATAA 2580  
Db 2521 TTAGTTAAAAACCTCTCTACTTAATTTTTTAAATATATAGAAATCTTTTCATGCAGGTTTCATAA 2580  
QY 2581 ACGGAGGCTTGAAGAAAGTGAAGTCTCGGAATACTTTCAACAGCTTGTAGATGCTGTGTC 2640  
Db 2581 ACGGAGGCTTGAAGAAAGTGAAGTCTCGGAATACTTTCAACAGCTTGTAGATGCTGTGTC 2640  
QY 2641 TCATTGTCACTGCAAGGGTGTTTACCACCGTGACCTAAGGTAAAGACGTTGTTTTGTTTT 2700  
Db 2641 TCATTGTCACTGCAAGGGTGTTTACCACCGTGACCTAAGGTAAAGTAAAGACGTTGTTTTGTTTT 2700  
QY 2701 ACCAATATTCCTCAGAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTTGTGTCGC 2760  
Db 2701 ACCAATATTCCTCAGAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTTGTGTCGC 2760  
QY 2761 TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATGGAAATCTGAAGGTTTT 2820  
Db 2761 TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATGGAAATCTGAAGGTTTT 2820  
QY 2821 CCGATTTCGGACTCAGTGCATTGCCTCAGGAAGTAAGTGTCTTATCTCTGCTTCAGCAG 2880

Db 2821 CGGATTTCCGACTCAGTGCATTGCCTCAGGAAGTAAGTGCTCTTATCTCTGCTTCAGCAG 2880  
QY 2881 TCTGCTTACGTGGTCAATAACTTGTATATATACTCAATCAGGGAGTAGAACTTCTGCGTAC 2940  
Db 2881 TCTGCTTACGTGGTCAATAACTTGTATATATACTCAATCAGGGAGTAGAACTTCTGCGTAC 2940  
QY 2941 CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTACTTAGTGGACAGGTTACGATGG 3000  
Db 2941 CACATGTGGAACCTCCGAACTATGTAGCTCCAGAGGTACTTAGTGGACAGGTTACGATGG 3000  
QY 3001 TTCAGCAGCTGATATTTGGTCTTTGCGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT 3060  
Db 3001 TTCAGCAGCTGATATTTGGTCTTTGCGGGGTTATTTCTTTTCGTTATATTTGGCTGGATATTT 3060  
QY 3061 ACCTTTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTCCG 3120  
Db 3061 ACCTTTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGTAACATATCTTTCCG 3120  
QY 3121 GAAGAAATCATGAATTCCTTGTCAATGGCTTTTGTCAAAACCGTTTTATTGATTTGGTTTTGC 3180  
Db 3121 GAAGAAATCATGAATTCCTTGTCAATGGCTTTTGTCAAAACCGTTTTATTGATTTGGTTTTGC 3180  
QY 3181 AATTTCAACCTTAGTTTGAGCTTTTACACATTGTTATTTACAGATAAAATGCAGCAGAGT 3240  
Db 3181 AATTTCAACCTTAGTTTGAGCTTTTACACATTGTTATTTACAGATAAAATGCAGCAGAGT 3240  
QY 3241 TTTCTTGTCCACCGTGGTTTTCCGCAGAAGTGAAGTTTTTAATACATAGGATACTTGACC 3300  
Db 3241 TTTCTTGTCCACCGTGGTTTTCCGCAGAAGTGAAGTTTTTAATACATAGGATACTTGACC 3300  
QY 3301 CCAATCCCAAAACAGTGAGTATTTTGTCTTGTCTCCTAGCTATCAGGTTTTTGGTGAT 3360  
Db 3301 CCAATCCCAAAACAGTGAGTATTTTGTCTTGTCTCCTAGCTATCAGGTTTTTGGTGAT 3360  
QY 3361 ATTTAAATGTTCTAGTAATATATCTGTTTATCTATTAATGTTTTCTCAATTAGACAGCGT 3420  
Db 3361 ATTTAAATGTTCTAGTAATATATCTGTTTATCTATTAATGTTTTCTCAATTAGACAGCGT 3420  
QY 3421 ATTTCAAAATTCAGGAAATCAAGAAAGATCCTTGGTTCAGATTAAATTTATGTGCCTATACGA 3480  
Db 3421 ATTTCAAAATTCAGGAAATCAAGAAAGATCCTTGGTTCAGATTAAATTTATGTGCCTATACGA 3480  
QY 3481 GCAAGGGAAGAAAGAAAGTGAATTTGGATGATFATTCGTGCAGTTTTTGTATGGAATTGAG 3540  
Db 3481 GCAAGGGAAGAAAGAAAGTGAATTTGGATGATFATTCGTGCAGTTTTTGTATGGAATTGAG 3540  
QY 3541 GTTTGTGCTTCGCCTTCAATATATATGCTCTTTGCTGGTCAATTCCAATTTAAATGTTAAGAT 3600  
Db 3541 GTTTGTGCTTCGCCTTCAATATATATGCTCTTTGCTGGTCAATTCCAATTTAAATGTTAAGAT 3600  
QY 3601 CTCTTAGGAACGTTTTGGATGACCAAGAAAGTGTGTCTACAGGATAGAACAAAAATAGT 3660  
Db 3601 CTCTTAGGAACGTTTTGGATGACCAAGAAAGTGTGTCTACAGGATAGAACAAAAATAGT 3660  
QY 3661 AGGCATGTGTTAGTTACCAAAACCTGTAAACTGCTTTTATTCAATTCGCCAAACCATAG 3720  
Db 3661 AGGCATGTGTTAGTTACCAAAACCTGTAAACTGCTTTTATTCAATTCGCCAAACCATAG 3720  
QY 3721 ACCTTAGGAAGACTTAGATTTTACAAGAGAAATCTCTATTCTCGACCAAAAACCTTAGACA 3780  
Db 3721 ACCTTAGGAAGACTTAGATTTTACAAGAGAAATCTCTATTCTCGACCAAAAACCTTAGACA 3780  
QY 3781 AAATCCAGAATACCCCTAGGGCTAATTACAAATGTTTCATGTACCTATCAATATATATCTCC 3840  
Db 3781 AAATCCAGAATACCCCTAGGGCTAATTACAAATGTTTCATGTACCTATCAATATATATCTCC 3840  
QY 3841 TGCTAAGAATGCTTAGCTTTGGTATAGCCTTAGATATATATAGATACCGTAATTTCTAAT 3900  
Db 3841 TGCTAAGAATGCTTAGCTTTGGTATAGCCTTAGATATATATAGATACCGTAATTTCTAAT 3900  
QY 3901 GCATATTTTAGAATGTTTTACTTAAACTCAGGCTTTGCTCTTCTTCTAAACCTTGACTTCAA 3960  
Db 3901 GCATATTTTAGAATGTTTTACTTAAACTCAGGCTTTGCTCTTCTTCTAAACCTTGACTTCAA 3960

QY 3961 TTGTTAAACTAAACCTCAGTATCTGCTTAGCTAAAGTTACTTTTACTTTTTCATT 4020  
Db 3961 TTGTTAAACTAAACCTCAGTATCTGCTTAGCTAAAGTTACTTTTACTTTTTCATT 4020  
QY 4021 AAGTTGACCTGTCAATTTGCACTTTGTTTCAAGGCGAGTTATGTAGCGAGAATGTAGAGAG 4080  
Db 4021 AAGTTGACCTGTCAATTTGCACTTTGTTTCAAGGCGAGTTATGTAGCGAGAATGTAGAGAG 4080  
QY 4081 AAATGATGAAGGCGCCCTGATGATGAATGCTTTGAGATGATTACCTTATCAAGGCTT 4140  
Db 4081 AAATGATGAAGGCGCCCTGATGATGAATGCTTTGAGATGATTACCTTATCAAGGCTT 4140  
QY 4141 AAATTTATCTGCACTATTGACAGGCGACAGGTAGTACCTGATTTTCTATTTACTGTTTCTAT 4200  
Db 4141 AAATTTATCTGCACTATTGACAGGCGACAGGTAGTACCTGATTTTCTATTTACTGTTTCTAT 4200  
QY 4201 AGAGATCTCCATTTTCGAATAAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT 4260  
Db 4201 AGAGATCTCCATTTTCGAATAAAAGAAATGTCGGTAGCATCTATTCTTCAGACTGCCCGTTT 4260  
QY 4261 TGACTGCCCTTATGATGCTGCTGTTCTTAGTTTGTATATAAATACTATAAGTTTCATTAGATGA 4320  
Db 4261 TGACTGCCCTTATGATGCTGCTGTTCTTAGTTTGTATATAAATACTATAAGTTTCATTAGATGA 4320  
QY 4321 TTGGTTGCATGGCATTTAGTAGATACAAAATGGAATCCAAAATGTTCCCTGCTATATTGATGGC 4380  
Db 4321 TTGGTTGCATGGCATTTAGTAGATACAAAATGGAATCCAAAATGTTCCCTGCTATATTGATGGC 4380  
QY 4381 TGATCCTTTTGATCTCGCAGGATTTTGTAAAGGCAAAACCCGTTTGTTCGAAAGGGA 4440  
Db 4381 TGATCCTTTTGATCTCGCAGGATTTTGTAAAGGCAAAACCCGTTTGTTCGAAAGGGA 4440  
QY 4441 ACCTAGTGAGATAAATGCTAACATTTGAGGCTGAGGAACTCAATGGGTTTAAAGTCTCA 4500  
Db 4441 ACCTAGTGAGATAAATGCTAACATTTGAGGCTGAGGAACTCAATGGGTTTAAAGTCTCA 4500  
QY 4501 TACAGAAACTTCAAGGTAAACGAATTCCTAGCATATTACATTTACAGAGATTATGCA 4560  
Db 4501 TACAGAAACTTCAAGGTAAACGAATTCCTAGCATATTACATTTACAGAGATTATGCA 4560  
QY 4561 TTATTTTAAAACTCTCAACTGTTAAACGCATGTTAGATAGATTGATAAGATTGACAAGG 4620  
Db 4561 TTATTTTAAAACTCTCAACTGTTAAACGCATGTTAGATAGATTGATAAGATTGACAAGG 4620  
QY 4621 AAACCTAGTTTATATCTCTGGGTTCAAAACGAAAGTCTAGTGTGAATTTATCATTTT 4680  
Db 4621 AAACCTAGTTTATATCTCTGGGTTCAAAACGAAAGTCTAGTGTGAATTTATCATTTT 4680  
QY 4681 AATGTTAGCAGAGAGTACAATTTGTTATGATTTGTTACGTTAGTGTGCTCAACAGACAAGG 4740  
Db 4681 AATGTTAGCAGAGAGTACAATTTGTTATGATTTGTTACGTTAGTGTGCTCAACAGACAAGG 4740  
QY 4741 CTGAGGGATTATCTTCGATCAAGGCGGACAGTTAGCTGTTGTTGATAGAGTAATTATT 4800  
Db 4741 CTGAGGGATTATCTTCGATCAAGGCGGACAGTTAGCTGTTGTTGATAGAGTAATTATT 4800  
QY 4801 GCTTGTGTGATTGTAATATAAGTTTGTCTTTCAGTTTAAAGGTTATCTACCAAAATTG 4860  
Db 4801 GCTTGTGTGATTGTAATATAAGTTTGTCTTTCAGTTTAAAGGTTATCTACCAAAATTG 4860  
QY 4861 AAATTAACCTACATGCAGATTACAGAGGTGGCACCATTGCTTTTTCATGGTAGCGTAAGA 4920  
Db 4861 AAATTAACCTACATGCAGATTACAGAGGTGGCACCATTGCTTTTTCATGGTAGCGTAAGA 4920  
QY 4921 AAGGCTGCTGGTGAACCTCTTGAATATCAAGGTTTAAATATATATATCCATTAACAAT 4980  
Db 4921 AAGGCTGCTGGTGAACCTCTTGAATATCAAGGTTTAAATATATATATCCATTAACAAT 4980  
QY 4981 AGTTGCATCATTTACTGTGTTGCGGATTAGAGTGATATTTGTTTGTGTTATCGCAGTTT 5040  
Db 4981 AGTTGCATCATTTACTGTGTTGCGGATTAGAGTGATATTTGTTTGTGTTATCGCAGTTT 5040

QY 5041 TACAAGAAGCTATGTTTGAAGAACTGGAATAACATAATATGGAGGCAACAGAAGGAATACCA 5100  
Db 5041 TACAAGAAGCTATGTTTGAAGAACTGGAATAACATAATATGGAGGCAACAGAAGGAATACCA 5100  
QY 5101 AAGTCAGAGATTCTCAGAACAAATCACGTTTGTATCCCAACTTAA 5144  
Db 5101 AAGTCAGAGATTCTCAGAACAAATCACGTTTGTATCCCAACTTAA 5144

RESULT 2  
US-09-938-842A-2784  
; Sequence 2784, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; TITLE OF INVENTION: SAME, AND METHODS OF USE  
; FILE REFERENCE: SRIPI300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 2784  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-2784

Query Match 8.1%; Score 414.6; DB 10; Length 2000;  
Best Local Similarity 87.8%; Pred. No. 3.8e-61;  
Matches 489; Conservative 0; Mismatches 59; Indels 9; Gaps 3;

QY 1 TTTTGGCGGAAATCTCGGGTTTACGTTTTCGGGGGAAATCTCGGTTTACGTTT 60  
Db 510 TTTTGGTGAGAAATTTACGGGTTTACGTTTTCGGGGGAAATTTACGGGTTTACGTTT 569  
QY 61 TGGCGGAAATCTCGGGTTTACGTTTTCGGAGAAATCACGGGTTTACGTTT 120  
Db 570 TGGCGGAAATTTATGGAATTACGTTTTCGGAGAAATCACGGGTTTAC-GTTTTGG 628  
QY 121 CGGAAATCACGGATTATGTTTTCGGTGGAAATTTACGAGTTTACGTTTCTCAAT 180  
Db 629 CGAGAAATCACGGGTTTATGTTTTCGGTGGAAATTTACGAGTTTATTTTCTCAAT 688  
QY 181 TTCATTGCTTGTATATTAAAGAAATTTGGAAATATTAATTTTAAATTTGTTTGA 240  
Db 689 TGCAATGCTTGTATATTAAAGAAATTTGGAAATATTAATTTTAAATTTGTTTGA 748  
QY 241 TGTGTTGTTAAACCTAAATGGCATTGGTTTACGAGATTGTTGTTTATTTCAATT 300  
Db 749 TGTGTTGTTAAACCTAAATGGCATTGGTTTACGAGATTGTTGTTTAAATTTCAAAT 808  
QY 301 TACAAATTTGATGGGTTAAATGGATAAACCAATGGAACCAATTAACCATTAACCCCAAC 360  
Db 809 TACAAATTTGATGGGTTAAATGGGTAACCAATTAACCAATTAACCATTAACCCCAAC 868  
QY 361 TCATTTTACTCATCAACCAATTTGACTCATCACTCACTTGAATTTGACTCATCACTCATTTGAG 420  
Db 869 TCATTTTACTCATCAACCAATTTGACTCATCACTCACTTGAATTTGACTCATCACTCATTTGAG 928  
QY 421 TCAAAATTTCAACTCATTTAGGTTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480  
Db 929 TAAAAAAATTT-----TAAGGTTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTT 981  
QY 481 TTGACCCATTTTGACACCCCTT-ACATATGATCATAGTTAAATAATCAAAATTTACTATTG 539

Db 982 TTGACCCATTTTGACACCCCTAACCGGTCAATTATATAATAATTTTATTATTAATG 1041  
QY 540 ATAACTTTTAAACGGAA 556  
Db 1042 TAGTTATTATTATTA 1058

RESULT 3  
US-09-938-842A-5327/c  
; Sequence 5327, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRI1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 5327  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-5327

Query Match 7.8%; Score 401.4; DB 10; Length 2000;  
Best Local Similarity 89.4%; Pred. No. 6.8e-59;  
Matches 455; Conservative 0; Mismatches 51; Indels 3; Gaps 2;  
QY 2 TTTTGGCGGAAAACTCGGGTTTACGTTTTCGTTTGGCGGAAAAATCTCGTTTACGTTT 61  
Db 1187 TTTTGGCGGAAAAATACGAGTTTACGTTTTCGTTTGGCGGAAAAATACGAGTTT 1128  
QY 62 GCGGAAAAATCTCGGGTTTACGTTTTCGTTTGGCGGAAAAATACGAGTTTTCGTTT 121  
Db 1127 GGCTGAAAAATACGAGTTTACGTTTTCGTTTGGCGGAAAAATATGAGTTTAC- 1069  
QY 122 GGGAAAAATCACGGATTTAAGTTTTCGTTTGGGAAAAATACGAGTTTACTTTCTCAAT 181  
Db 1068 GGGAAAAATCACGGGTTTATCTTTTAG- -TGAAAAATACGAGTTTACTTTATCAAT 1011  
QY 182 TCATTGCTGTATATTTAAGAAATTTGAAAAATATTAATTTTATTAATTTGGTTAGAT 241  
Db 1010 TCATCGCTGTATATTTAAGAAATTTGAAAAATATTAATTTTATTAATTTGGTTAAAT 951  
QY 242 GTGTTGGTTAAACCTAAATTTGGCATTTGGTTTAGAGATTTTAGTTGGTTTATCAATTT 301  
Db 950 GTGTTGGTTAAACCTAAATTTGGCATTTGGTTTAGAAATTTTAGTTGGTTTAAATTT 891  
QY 302 ACAAAATTTGATGGTTAATTTGGATAAACCATGGAACCAATTAACCATTAACCTAACT 361  
Db 890 ACAAACTTTGATGGTTAATTTGGTAACCATTTGAATCAATTAACCATTAACCAACT 831  
QY 362 CATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCACTCACTCACTCACT 421  
Db 830 CATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCACTCACTCACTCACT 771  
QY 422 CAAAAATTTCAACTCATTAGGTTTCATGGTTTGGTTGAGTTGAGTTGAGTTGAGTT 481  
Db 770 CAAAAATTTCAACTCATTAGAGTTTCATGGTTTGAATTTGAGTTGAGTTGAGTTGAGTT 711  
QY 482 TGACCCATTTTGACACCCCTACATGAT 510

Db 710 TAACCCATTTTGACATCTCTACCTGCGTT 682  
RESULT 4  
US-09-938-842A-4657  
; Sequence 4657, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SCRI1300-3  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/300,111  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 5379  
; SEQ ID NO 4657  
; LENGTH: 2000  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-938-842A-4657

Query Match 7.2%; Score 371.2; DB 10; Length 2000;  
Best Local Similarity 95.3%; Pred. No. 9.8e-54;  
Matches 404; Conservative 0; Mismatches 18; Indels 2; Gaps 2;  
QY 47 CGTGTTCACGTTTTCGCGGAAAAATCTCGGGTTTACGTTTTCGCGGAAAAATCACGGG 106  
Db 748 CGGTTTACGTTTTCGCGGAAAAATCTCGGGTTTACGTTTTCGCGGAAAAATCATGGA 807  
QY 107 TTACTTTTTCGCGGAAAAATCACGGATTTATGTTTTCGTTGGAAAAATACGAGTT 166  
Db 808 TTAC-GTTTTTCGCGGAAAAATCACGGATTTACG-TTTTTTCGCGGAAAAATACGGTT 865  
QY 167 TACTTTTCTCAATTTTCATGCTTGTATATTTTAAGAAATTTGGAAAAATATTAATTTAT 226  
Db 866 TACTTTTCTCAATTTTCATGCTTGTATATTTTAAGAAATTTGGAAAAATATTAATTTAT 925  
QY 227 TAAATTGGTTTAGATGTTGTTTAAACCTAAATTCGATTTGGCATTGGTTTAGAGATTTAGTTG 286  
Db 926 TAAATTGGTTTAGATGTTGTTTAACTTAAATTCGATTTGGCATTGGTTTAGAGATTTAGTTG 985  
QY 287 GTTTTATTCATTTTACAAAAATTTGATGGTTTAAATTCGATAAACCATGGAACCACTAAC 346  
Db 986 GTTTAATTCATTTTACAAAAATTTGATGGTTTAAATTCGATAAACCATGGAACCACTAAC 1045  
QY 347 CATTACAACTTAATCTATTTTACTCATCAACCAATTCGATCACTCACTCACTCACTCACT 406  
Db 1046 CATTACAACTTAATCTATTTTACTCATCAACCAATTCGATCACTCACTCACTCACTCACT 1105  
QY 407 TCAACTCATTTGAGTCAAAAATTTCAACTCATTTAGGTTTCATGGTTGAGTTGAGTTGAG 466  
Db 1106 TCAACTCATTTGAGTCAAAAATTTCAACTCATTTAGGTTTCATGGTTGAGTTGAGTTGAG 1165  
QY 467 TTGA 470  
Db 1166 TTGA 1169

RESULT 5  
US-09-938-842A-5099/c  
; Sequence 5099, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff



```

; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5099
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5099

```

Query Match 3.0%; Score 152.4; DB 10; Length 2000;  
Best Local Similarity 91.0%;  
Pred. No. 2.2e-16;  
Matches 162; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 338 ACCATTAACCATTAACAACCTAACTCATTTTACTCATCAAACCAATTGACTCATCACTCA 397

Db 1557 ACCATTAACCATTAACAACCAACTAATTTTACTCATCAAACCAATTGACTCATCACTCA 1498

Qy 398 TTTGACTCATCAACTCATTTGAGTCAAAAATTTCAACTCATTAGGTTTCATGGTTTCAGT 457  
 |||||  
 Db 1497 TTTGACCCATCAACTCATTTGAGTCAAAAATTTTGAACACTCATTAGAGTTTCATGGTTTCAGT 1438  
 |||||

[illegible]

```

RESULT 6
US-09-878-574-3933
; Sequence 3933, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Oth
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3933
; LENGTH: 404
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-003-Q1-B1-G6
US-09-878-574-3933

```

Query Match	1.9%;	Score 97.6;	DB 10;	Length 404;
Best Local Similarity	64.7%;	Pred. No. 2.6e-07;		
Matches 145;	Conservative	0;	Mismatches 79;	Indels 0;
				Gaps 0;

Qy 2916 ATCAGGGAGTAGAACTTCTCGGTACCACATGTGGAACTCCGAACTATGTAGTCTCCACAGG 2975

Db 78 AACAGGAGGATGAACCTTCTCCGCACCTGTTGTGGAACTCCAAATTATGTTGTTCTCTGAGG 137

Qy	2976	TACTTAGTGGACAGAGGTTACGATGGTTCAGCAGCTGATAATTGGTCTTGCGGGGTATTTC	3035
Db	138	TGCTTAATGATAGAGGTTATGTTGGTCTACATCTGATATCTGGTCCCTGTGGAGTCATTC	197

[illegible]

RESULT 7

US-10-292-408-8

; Sequence 8, Application US/10292408

; Publication No. US20030182692A1

; GENERAL INFORMATION:

; APPLICANT: VAN THIELEN, NOCHA

; APPLICANT: DA COSTA E SILVA, OSWALDO

; APPLICANT: CHEN, RUOYING

; TITLE OF INVENTION: PROTEIN KINASE STRESS-RELATED POLYPEPTIDES AND METHODS

; TITLE OF INVENTION: OF USE IN PLANTS

; FILE REFERENCE: 16313-0178

; CURRENT APPLICATION NUMBER: US/10/292,408

; CURRENT FILING DATE: 2002-11-12

; PRIOR APPLICATION NUMBER: 60/346,096

; PRIOR FILING DATE: 2001-11-09

; NUMBER OF SEQ ID NOS: 51

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 8

; LENGTH: 1468

; TYPE: DNA

; ORGANISM: Physcomitrella patens

US-10-292-408-8

Query Match	1.9%;	Score 95.6;	DB 12;	Length 1468;
Best Local Similarity	67.7%;	Pred. No. 9.8e-07;		
Matches 134;	Conservative	0;	Mismatches 64;	Indels 0;
				Gaps 0;

[illegible]

```

RESULT 8
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeuten
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated ge
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)

```



US-10-312-841-1

Query Match	1.8%;	Score 91.6;	DB 12;	Length 3673778;
Best Local Similarity	43.2%;	Pred. No. 0.00012;		
Matches 546;	Conservative	0;	Mismatches 709;	Indels 9;

QY	181	TTCAATTGCTTGTATATATTTAAAGAAATTTGGAAAAATATAATTTTATTAAATTTGGTTTTAGA	240
D6	1713092	TTAATAATAATATATATATATAAAAATAATATAATATAAACGTATTAATTTATATTTTG	1713151
QY	241	TGTGTTGGTTAAACCTFAAAATGGCAATGGTTTAGAGATTTTAGTTGGTTTTATTCCAATTT	300
D6	1713152	TGTGTTGTATATATATATTTTATATATATATATTTTATATATATATTTTATATATATATGTAATTT	1713211
QY	301	TACAAAAATTTGATGGGTTAAATTGGATAAAACCATTGGAACCATTAACCATTAACAACCTAAC	360
D6	1713212	TATATATATTTTATATATATATTTTATATATTTTATATATTTTATATATATTTTATTTAT	1713271
QY	361	TCATTTTACTCATCAACCAAAATTGACTCATCAACTCATTTGACTCATCAACTCATTTTGAG	420
D6	1713272	ATATTTTATATATTTTATATATATGTTTTATATATTTTATATATATATTTTATATATATA	1713331
QY	421	TCAAAAAATTTCAACCTCATTAGGGTTCAATGGGTTGAGTTGAGTTGAGTTGACCCCATGAAT	480
D6	1713332	TTTTATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATATATAT	1713391
QY	481	TTGACCCATTTTGACACCCCTACATATGATCATAAAGTTAAATAATCAAATAATTACTATTGA	540
D6	1713392	TTTATATATATTTTATATATATATTTTATATATATATATTTTATATATATATTTTATATATAT	1713451
QY	541	TAACTTTTTAAACGGAATTGTTTTTGTAAAGTTTCATTTACGTTTATTTTATATAAAAAACATA	600
D6	1713452	TTATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATATATATAT	1713511
QY	601	ATGCAAAAGTACTAAATGTPATAGTTATTTTTTATTTTAAATAATGC AAAAATATTACTGTAA	660
D6	1713512	TTTATATATATAAATTTATATATATATATTTTATATATATATTTTATATATATATTTTATATA	1713571
QY	661	TACTTTTTTCATTCCTTATCAATATTTTTTCTTTATATTTCAATTAGCCACCAACCTACATT	720
D6	1713572	TTTTATATATTTTATATATATTTTATATATTTTATATATATATTTTATATATATTTTATATA	1713631
QY	721	TCATATTTCTCTCACATTCATTATTTTCTTGCTCTTGTTATTCCTTCCTATCATCTTCAT	780
D6	1713632	TTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATA	1713691
QY	781	TGCCAAATTTTTCATTGACATTGTCATCGTTTACTTTGTATGTATGATTTTGAACATTTAA	840
D6	1713692	TATATTTTATATATTTTATATATATAAATAATATATATATATATATATATATATATATATTA	1713751
QY	841	TGTGAATCCCCAACCGACTATAGGACTATCAAGAGTTTTCAAACTTTTAAAAAAGATCT	900
D6	1713752	TATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATTTTATATATATTT	1713811
QY	901	TGAACCTTTTAAAGCAAAATCCC AACCGACTATAGAACCATAGTCCCTCTCTGTTGATGA	960
D6	1713812	TATATATTTTATATATATTTTATTTTATATATTTTATATATATATATATATATATATATAT	1713871
QY	961	AGCTCTTCTCGTGCCGGGAAAAATCTAGGCCATAAAAAGCCTCTTCAACATCACCTAGTAT	1020
D6	1713872	ATATTTTATATATATTTTATATATATTTTATATATATATATATATATTTTATATATATTTTAT	1713931
QY	1021	AATTGACCGTGACCATCTTTTGTGACCATTTGCTTTTGGAATGAACCGTCGATAAACCGTGT -	1079
D6	1713932	ATATATATTTTATATATATTTTATATATATATATATTTTATATATATATTTTATATATATTTTAT	1713991
QY	1080	--TATCACTTACGCCAAAATTTTCCCTAGTGTGTTTGTGTTTCCAAACTCTCAACGAATCCTTAT	1137
D6	1713992	TATATATTTTATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTAT	1714051
QY	1138	CGAACTTTTATATATATCACTTTGTAGCAT-----TGGAAAGTATCTTTGTATGCTTTTG	1191
D6	1714052	ATACATTTTATATATATTTTATATATATATTTTATATATATATTTTATATATATATTTTATATA	1714111

QY	1192	TCTTAAACCTTAGACATCCTTGTTCTCTCTGTTTCTGTTGAACTTGCTTGACTTAAATGAA	1251
Db	1714112	TATATATTTTATATATATTTTATATATATATTTTATATATATATATTTTAT	1714171
QY	1252	GTTAAAAAATTTGAGTTAAAAAATGAAAAATTTTACTAAATTTGAGTTCGATTGAATCATAGT	1311
Db	1714172	ATATATTTTATATATATTTTATATATATATTTTATATATATTTTATATATATATATATAT	1714231
QY	1312	CTAGATAAATTTGAAAAAATTAATAAAAAATTTGAAAAATCTATATGTTATATTTTATAA	1371
Db	1714232	TTATATATTTTATATATATATTTTATATATATATTTTATATATATTTTATATATATTTTATATA	1714291
QY	1372	AATAATTTTACTAAATTGATAAGTGGATATTAGATTAGTTTTTTTTCTTTTTTAAATTTT	1431
Db	1714292	TATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATATTT	1714351
QY	1432	GAAA	1435
Db	1714352	TATA	1714355

## RESULT 9

```

US-09-770-149-51
; Sequence 51, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickner, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 742
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-149-51

```

```

Query Match      1.7%; Score 88; DB 9; Length 742;
Best Local Similarity 62.1%; Pred. NO. 1.5e-05;
Matches 139; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```

Qy	1947	TCTGAATAATCTGATTGATAGCTTTGTGGTCAATGGATCAGATAAAAGTTTGTAAAGAAATG	20066
Db	57	TCTAAGTTTTCTGAAGGAGTGTATATTTGTTTGGTGGTGTAGAGAAATGAATCGGAGA	116
Qy	2007	ACAAAGAAAATGAGAAAGTGGCGAAGTACGAGGTTTGGTTCGCACAATAGGTGAAGGAACC	20666
Db	117	CAGCAAGTGAAACGTAGAGTAGGTAAATATGAAGTTTGGAAAGAACAAATTTGGAGAGGAACG	176



```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 72
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-72

Query Match      1.6%; Score 82.2; DB 10; Length 1326;
Best Local Similarity 71.5%; Pred. No. 0.00018;
Matches 108; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 2020 GAAGAGTGGGCAAGTACGAGGTTGGTCGCACAAATAGGTGAAGGAACCTTTGCTAAGGTTA 2079
Db 26 GTAGAGTAGGTAATATGAAGTTGGAAGAACAAATGGAGAAGGAACGTTTGCTAAAAGTTA 85

QY 2080 AGTTTGCAGGAACACAGACACACTGGTGATAATGTAGCCCATCAAAATTTATGGCTAAGAGTA 2139
Db 86 AGTTTGCTAGAAACTCTGAAACTGGAGAACCTGTGCTCTCAAGATTCCTTGATAAAGAGA 145

QY 2140 CAATACTTAAGAACACAGAATGGTTGATCAGGT 2170
Db 146 AAGTCTCAAGCATAAATAATGGCTGAACAGAT 176

RESULT 12
US-10-311-455-2128
; Sequence 2128, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2128
; LENGTH: 15548
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2128

Query Match      1.6%; Score 81.8; DB 12; Length 15548;
Best Local Similarity 43.8%; Pred. No. 0.00058;
Matches 598; Conservative 0; Mismatches 752; Indels 15; Gaps 5;

QY 142 TTTTGTGCGAAAAAATTACGAGTTTACCTTTCTCAATTCATTCGCTGTATATTAAAG 201
Db 11398 TTCGTTGTATAATAATTAAGTAATTAATAGTAATTATATAATTAATAATGAATGTTTGA 11457

QY 202 AAATTTGGAAAAATATTAATTTTAAATTTAGATGTGTGTTAAACCTAAAT 261
Db 11458 AAGAATAATGTAATGAATATATATGATTTTATTATATATATTTTATTTTATTTTAT 11517

QY 262 GGCATTGGTTTAGAGATTTTAGTTGGTTTATTCAAATTTTACAAAATTTGATGGGTTAAT 321
Db 11518 TTATATATTTTATTTTATATATTTTATATATTTTATTTTATTTTATTTTATTTTAT 11577

QY 322 TGGATAAACCATGGAAACCAATTAACCATTAACAACCTAATCTATTTTACTCATCAAAACCAA 381
Db 11578 TTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 11637
```

```

QY 382 TTGACTCATCACTCATTTTGACTCATCACTCACTCATTTGAGTCAAAAATTTCAACTCATTAG 441
Db 11638 TATATTTTATTTTATATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 11694

QY 442 GGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 501
Db 11695 TTTTATATATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11754

QY 502 ACATATGATCATAAAGTTAAATAATCAAAAAATTTACTAATGATAAATTTTAAACGGAATTTGTT 561
Db 11755 TTATTTTATATTTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11814

QY 562 TTGTAAGTTTCATTTACGTTTATTTATATAAAAAAACATAATGCAAAAGTACTAATGTATA 621
Db 11815 ATATTTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11874

QY 622 GTTATTTTATTTTAAATAATGCAAAAAATTTACTGTAATACTTTTTCATTTCTTATCAAT 681
Db 11875 TTTATTTTATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11934

QY 682 ATTTTTCCTTATATTCAATTAGCCCAACCACTTACATTTTCATATTTTCTTTCACATTCA 741
Db 11935 TTTTATATATGTTATTTTATATGTTATATATATTTTATTTTATATATGTTATATATATTTTATA 11994

QY 742 TTATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 796
Db 11995 TTTTATATATGTTATATATATTTTATTTTATATATTTTATTTTATATGTTATTTTATATATG 12054

QY 797 ACATTGTCATCGTTACTTTTGTATGATGATGATTTTGAACATTTTAAATGTAATCCCAACCGA 856
Db 12055 TTATTTTATATGTTATTTTATATGTTATTTTATATGTTATTTTATATGTTATTTTATTTT 12114

QY 857 CTATAGGACTATCAAGAAGTTTTCAAAATTTTAAAAAAGATCTTGAACCTTTTAAAGCAA 916
Db 12115 ATATGTTATTTTATATGTTATATATATATATGTTATTTTATAGTATATATATATATATGTTAT 12174

QY 917 ATCCCAACCGACTATAGA - ACCATAGTCTCTTCTTCTTCTTGTGTGATGAAGCTCTTCTCGTGCC 975
Db 12175 ATATATATGTTATATATATATATGTTATATATATATATGTTATATATATATATATATATATAT 12234

QY 976 GCGGAAAATCTAGGCCATAAAAGCCTCTTCAACATCACCTAGTATATTGACCGTGACCAT 1035
Db 12235 ATGTTATATATATATATGTTATATATATATATATGTTATATATATATATATATATATATATAT 12294

QY 1036 CTTTTTGACCATTTGCTTTTGTGAATGAACCGTCGATAAACCGTGTATCACTTACGCCAAA 1095
Db 12295 TATATATATATGTTATATATATATATATGTTATATATATATATATATATATATATATATATAT 12354

QY 1096 TTTTTCCTTAGTGTGTTTCCAAAATCTCACGAATCCCTTATCGAACTTTTATATATAT - 1154
Db 12355 TATATATGTTATATATATATATATGTTATATATATATATATGTTATGTTATATATATATATA 12414

QY 1155 ---CACTTTGTAGCATTTGGAAGTATCTTTGTATGCTTTGTCTTAAACCTTAGACATCCT 1210
Db 12415 TATATTTTATGTTATATATGTTATATATATATATATATTTTATGTTATATATTTTATATATATA 12474

QY 1211 TGTTCTCTTGGTTTTTTTGAACCTTGCTTGACITTAAGAAAGTTAAATAATTTGTAGTTAA 1270
Db 12475 TTTTATGTTATATATTTTATATATATATATATTTTATGTTATATATTTTATATATATATATTTT 12534

QY 1271 AAATAGAAAAATTTTACTAATTTGAGTTCGATTAAATCATAGTCTAGATA - ATTTGAAAAAA 1329
Db 12535 ATGTTATATATTTTATATATATATATATTTTATGTTATATATTTTATATATATATATTTTATGT 12594

QY 1330 ATTAAAAATAAATTTTGAATAACTATATATGTTTTTATTTTAAAAATAATTTACTAAATGA 1389
Db 12595 ATATATGTTATATATATATATATATATATATTTTCGTTATATATATATTTTATATATATATAT 12654

QY 1390 TAAGTGATATTAGATTAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1449
Db 12655 TATATATTTTATTTTCGTTATATATATATTTTATATATATATATATATATATATATATTTTCGTTATATA 12714
```



QY 1450 AAATTGAAGATTCAAAATGCAATGTTAGTGTCTTAAAGTTTAAAT 1494  
Db 12715 TATATTTATATATATATATATATATTTCTGATATATATAT 12759

RESULT 13  
US-10-311-455-157  
; Sequence 157, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 157  
; LENGTH: 15387  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-157

Query Match 1.6%; Score 80.4; DB 12; Length 15387;  
Best Local Similarity 44.1%; Pred. No. 0.001;  
Matches 572; Conservative 0; Mismatches 706; Indels 18; Gaps 5;  
QY 136 TTTATGTTTTTTGGTGGAAAAATACGAGTT--TACTTTTCTCAATTTCAATGCTTGT 192  
Db 9779 TTTGTTTTTTAGTTTAAAGATCGAAATTTGGGTGTTATTTTGTGATTTTTTTTTT 9838  
QY 193 ATATTAAGAAATTTGGAAAAATATTAATTTTATTAATGTTTGTAGATGTTGGTTAA 252  
Db 9839 ACGTTTTTATAGTTTATAGATATTAAGTTTGTGAAATTTTATTTTATTTTGTGTTA 9898  
QY 253 ACCTAAATGGCATTGGTTTAGAGATTTT-AGTTGGTTTTTATTAATCAATTTACAAAATTTG 311  
Db 9899 TTTTATGTTTATTTTATAGATATTAAGTTTGTGAAATTTTATTTTATTTTATTTTGT 9958  
QY 312 ATGGGTTAATTTGGATAAACCAATGGAACCAATTAACCAATTAACCAATTAACCAATTTACTC 371  
Db 9959 TTTGTTTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTT 10018  
QY 372 ATCAAAACCAATGACTCATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTC 431  
Db 10019 TGTGTTTTTTTATGTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 10078  
QY 432 AACTCATTTAGGTTTCATGGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTGAGTTG 491  
Db 10079 TTTTGTGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTAT 10138  
QY 492 TGACACCCCTACATATGATCATATGATTAATTAATCAAAATTTACTATTTGATAACTTTTAA 551  
Db 10139 TATTTTGTGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTATTTTATGTTAT 10198  
QY 552 CGGAATTGTTTGTAAAGTTTCATTTTACGTTTATTTATATATAAAAAACATAATGCAAAAGTA 611  
Db 10199 TTTTATTTTGTGTTGTTTATTTTATGTTATTTTATTTATTTATTTATTTATTTATTTATTTAT 10258  
QY 612 CTAATGTATAGTTATTTTATTTTAAATGCAAAATTAATGCTGTAATCTTTTCTCAT 671  
Db 10259 ATTTTATTTTGTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 10318

QY 672 TCTTATCAATATTTTCTTATATTTCAATTAGCCACCAACACCTACATTTTCTATTTCTC 731  
Db 10319 TTTAAATTTTATTTTGTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10377  
QY 732 TTCACATTTCAATTTTCTTCTGCTCTTGTATTTCTTCTTATCATCTTCTCAATTTT 791  
Db 10378 ---AAATTTTATTTTATTTTGTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10434  
QY 792 CATTGACATTTGTCATCGTTACTTTTGTATGATGATGATTTTGAACATTTTAAATGTAATCCCA 851  
Db 10435 TTTGTAAATTTTATTTTATTTTGTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10488  
QY 852 ACCGACTATAGGACTATCAAGAAGTTTTCAAACTTTTAAAGATCTTGAACCTTTAA 911  
Db 10489 --GATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 10546  
QY 912 AGCAAAATCCCAACCGACTATAGAACCATAGTCTTCTTCTTGTGATGAGAGCTCTTCTCG 971  
Db 10547 TATTAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10606  
QY 972 TGCCGGCGAAAAATCTAGGCCATAAAAGCCTCTTCAACATCACCTAGTATTTGACCGTGA 1031  
Db 10607 TTTTATAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10666  
QY 1032 CCATCTTTTGTACCATTTGTTGTAAGTAAACCGTGTGATGAGAGCTCTTCTCG 1091  
Db 10667 TTTTATAGATATTAAGTTTGTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10726  
QY 1092 CAAATTTTCTTCTAGTGTGTT- GTTCCAAACCTCAGGAATCCTTATCGAACTTTTAA 1149  
Db 10727 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10786  
QY 1150 TATATCACTTTGTAGCATTTGAAAGTATCTTTGTATGCTTTGTCTTAAACCTAGACATCC 1209  
Db 10787 TTTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10846  
QY 1210 TTTGTTCTTTGTTTGTGAACTTGTCTTGTGACTTAAATGAAAGTTTAAATTTTGTAGTTA 1269  
Db 10847 TTTATTTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10906  
QY 1270 AAAATAGAAAAATTTTACTAATTTGAGTTGCTGATTAATCATAGTCTAGATAATTTGAAAAAA 1329  
Db 10907 TTTTATTTGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 10966  
QY 1330 ATTAATAATAAATTTTGAAAAATCTATATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1389  
Db 10967 TTTATTTTATTTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11026  
QY 1390 TAAGTGATATTTAGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1425  
Db 11027 TGTGTTGTTTATTTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 11062

RESULT 14  
US-10-312-841-2/c  
; Sequence 2, Application US/10312841  
; Publication No. US20030186277A1  
; GENERAL INFORMATION:  
; APPLICANT: Epigenomics AG  
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC  
; FILE REFERENCE: E01/1208/WO  
; CURRENT APPLICATION NUMBER: US/10/312,841  
; CURRENT FILING DATE: 2002-12-30  
; NUMBER OF SEQ ID NOS: 2  
; SEQ ID NO 2  
; LENGTH: 367378  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
; NAME/KEY: unsure  
; LOCATION: (379615)



US-10-312-841-2

```
Query Match      1.5%; Score 77.4; DB 12; Length 3673778;
Best Local Similarity 44.2%; Pred. No. 0.031;
Matches 420; Conservative 0; Mismatches 521; Indels 10; Gaps 2;

QY      586 ATATAAAAAACATAATGCAAAAGTACTAATGTATAGTATATTTTATTTTAAATAATGCA 645
Db      1400891 AAAAAACGAATTTAAATACGAACCTTAATACAATATAACAATATAAATAAACAATAAT 1400832

QY      646 AAAATATTACTGTAAATCTTTTTCATTCTTATCAATATTTTTCTTATATCAATTAGCC 705
Db      1400831 AAAAAAATTAAAAATATAAATAATTTAACTTAAACTTTTATTTTAAATATAATATC 1400772

QY      706 ACCAACACCTACATTTTCATATTTCTCTTCACATTCATATATTTCTTGTCTGTATTCCT 765
Db      1400771 TAAACCTATATTTATATAAAAAATCGCATTTATTAACACTATAATTCCTATATAAAATCTTT 1400712

QY      766 TCTTATCATCTTCATTCGCCAATTTTTCATTTGACATGTTCATCGTTACTTTGTATGTATGA 825
Db      1400711 TTTAAACTAATTCAAATTTTCTTTTATTTTAAATTTTATCTCTTTTAAATAAAAAATATAA 1400652

QY      826 TTTTGAACATTTTAATGTGAATCCCAACCGACTATAGGACTATCAAGAAGTTTTTCAAAC 885
Db      1400651 TCTTTAAACTAATACTAAAAATTTTAAATAAATACGCTCTATATTAAAAAACTTTTACAAT 1400592

QY      886 TTTTAAAAAAGATCTTGAACCTTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCCT 945
Db      1400591 TTTTAAAAAACAAATTAATATTTTAAAAAAAACCTTATATTTTATTTAAATATTTTAA 1400532

QY      946 CTTCCTTGTGTGAAGCTCTTCTCGTGGCGGCAAAATCTAGGCCATAAAGCCTCTTC 1005
Db      1400531 TAAAAAACTAAATAATATCTTTTAACTTTTAAATAAATACGTTTACACAGAAATTTTTTA 1400472

QY      1006 AACATCACCTAGTATA-----TTGACCGTGACCATCTTTTGGACCATGTCTTTGTGAA 1058
Db      1400471 CAATTATCATTTTAAAAACTTATTTAAATCTTTAAAAACAAAATTAACAACCTTTTATA 1400412

QY      1059 TGAACCGTCGATAAACCGTGTATCACCTTACGC---CAAATTTTCCCTAGTGTGTTT 1115
Db      1400411 TAAATTTTATAAATCTTTTATAAATCTTTTACAAACAAATTTTAAACATATCTTAACTTT 1400352

QY      1116 CCAAACTCTACGAACTCTTATCGAACTTTTATATATATCATCTTGTAGCATTTGGAAGT 1175
Db      1400351 TTATATTTTATAAATTTTTTTACTAAAAAATACATTTTATAAATTTTAAATTTTACT 1400292

QY      1176 ATCTTTGTATGCTTGTCTTAAACTTTAGACATCTTGTCTCTTGGTTTTTTTGAACCTT 1235
Db      1400291 TTTTTHATAATTTTAAATTTTATCTTAATCTTTTCTTTTACCTTTTAAATATTTTAAATAT 1400232

QY      1236 GCTTGACTTAAATGAAGTTAAAAATTTGTAGTTAAAAATAGAAAATTTTACTAATTTGAG 1295
Db      1400231 ATAAATAATTAAATAAATATTAAATAACAATAAATATATATACATATTTTAAATTTTAAAT 1400172

QY      1296 TTCGATTAAATCATAGTCTAGATAAATTTGAAAAAAATTAATAAATTTTGAATACTAT 1355
Db      1400171 TTAAAAATATATTTAAACATCTATTTACCAAAACTAACTAAATTCCTTACGATTA 1400112

QY      1356 ATGTTTATTTTAAAAAATAATTTACTAAATTTGATAAGTGATATAGATTAGTTTTTTTT 1415
Db      1400111 ACACCTTAAAAAATAATATATACCTATAAATACTAAATAATCTAAACATTAATAAATAAT 1400052

QY      1416 TTCCTTTTAAATTTTGAACCTCACAATTTATTAATTTGAAGATTCAAAATGCAATGT 1475
Db      1400051 TATTTAACCAACCTCTATATAAATTTAAAAATTTTAAATAAATTTCTCCAATTTTAAATAA 1399992

QY      1476 TAGTTTTTAAAAAGTTTAAATCACCCCAACGGATAATTGACCCCAACGACTAA 1526
Db      1399991 AATAATCGATATAATTAATAAATTAATCAACAACAATAATATCATAAACCTAA 1399941
```

```
; Sequence 1610, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1610
; LENGTH: 7306
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1610
```

```
Query Match      1.5%; Score 76.6; DB 12; Length 7306;
Best Local Similarity 45.7%; Pred. No. 0.0033;
Matches 345; Conservative 0; Mismatches 404; Indels 6; Gaps 2;

QY      95  GAAATACACGGGTTTACTTTTTTTTGGCGGAAAAATCACGGATTTATGTTTTTGGTGGA 154
Db      5618  GAGTATCGTGTATTGTATTTTACGTTGGGTAACGGAGTGAGATTTATATGGAATAA 5677

QY      155  AAATTACGAGTTTACTTTTTCTCAATTTTCATTCGCTGTATATTTAAGAAAAATTTGAAAAA 214
Db      5678  AAATAAGAGGTTAGGATGTTTGGTGTATTGTTATTGTTTTATATATTTTGAAGGAT 5737

QY      215  TATTAAATTTTATAAATGGTTTAGATGTGTGGTTAAACCTTAAATGGCATTTGGTTTAG 274
Db      5738  TTGTTAAATGTTAAAGAAATATAAGGAAGGTAAGAGGTGAAAGAGAGAAATGAAATAT 5797

QY      275  AGATTTTAGTGGTTTTTATTCAATTT---TACAAAAATTTGATGGGTTAATTTGGATAAAC 330
Db      5798  TATTGTTGAAGATGATATTATTTTATATAGAAAAATTTGTAGAAATTAATGGTAAGT 5857

QY      331  CATGAAACCATTAACCATTAACCTTAACCTCAATTTTACTCATCAAACTTGAATTTGACTCAT 390
Db      5858  TATTAGAAATTTAAGAGAAATTTAGTGAGGTTGTTGTATTATGTTAAATTTTAAATAT 5917

QY      391  CAACCTCAATTTGACTCATCAACTCATTTT--GAGTCAAAAATTTCAACTCATTTAGGGTTTCA 448
Db      5918  TGATAGTATTTTGTAAATATTTTAAATAGTTATAAATAATAGTATAAAGTAGTATTA 5977

QY      449  GGGTTGAGTTGAGTTGAGTTGACCCCATGAATTTTGACCCATTTTGACACCCCTACATATG 508
Db      5978  AAAATATTAAATTTATTAGGAAATAATTTTATAGAGAAAAATTTAGTTTGTAAAGGATA 6037

QY      509  ATCATAAGTTAATAATCAAAAATTAATTTAGATAACTTTTAAACGGAATTTGTTTGTGAAG 568
Db      6038  AATAGTGGTAATGTACGTTATGTTTATAGAGATTATATTTTAGTTTAGTAAAGATATTA 6097

QY      569  TTTCAATTTACGTTTATTTATATAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTT 628
Db      6098  TTTTTTAAATTTTATTATAAAATTAATGTATGTGAATTAATAATTTTATTGGAATTT 6157

QY      629  TTATTTTAAATGCAAAAAATTAATCTGTAATACTTTTTCATTTCTTATCAATATTTTT 688
Db      6158  TTATTAGGAAGGTAATAAATTTTTTTTTTTTTTTTTTTTTTTTATTTTATTTATTTATTT 6217

QY      689  CTTATATTCAATTTAGCCCAACACCTACATTTTCATATTTCTCTTCACATTCATTTT 748
Db      6218  TATTATTATTATTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 6277
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 01:20:12 ; Search time 265 Seconds  
(without alignments)  
8567.830 Million cell updates/sec

Title: US-09-824-735-1  
Perfect score: 5144  
Sequence: 1 tttttgcggaataatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	79.2	1.5	299	4	US-09-313-294A-7428
2	66.6	1.3	19124	2	US-08-487-826B-13
3	65.4	1.3	658	3	US-08-998-416-186
C 4	63.8	1.2	6243	2	US-09-056-075-1
C 5	63.4	1.2	19124	2	US-08-487-826B-13
C 6	62.6	1.2	7218	1	US-08-232-463-14
7	60.6	1.2	5852	1	US-07-867-106-2
8	58.4	1.1	7218	1	US-08-232-463-14
9	57.8	1.1	53332	4	US-09-801-861-3
C 10	57.2	1.1	6265	4	US-09-129-112-3
C 11	57	1.1	5852	1	US-07-867-106-2
C 12	57	1.1	6152	3	US-08-973-462-1
13	57	1.1	20674	4	US-09-641-638-651
C 14	56.8	1.1	6124	4	US-08-213-419B-3
C 15	56.8	1.1	20674	4	US-09-641-638-651
C 16	56.8	1.1	1664976	4	US-08-916-421B-1
17	56.6	1.1	1250	3	US-08-688-988-1
18	56.6	1.1	1647	3	US-09-101-146-44
C 19	56.6	1.1	2287	3	US-08-845-258-8
C 20	56.6	1.1	2287	3	US-08-990-571-8
C 21	56.6	1.1	2287	4	US-08-723-142A-8
C 22	56.6	1.1	2287	4	US-09-528-784A-8
C 23	56.6	1.1	2287	4	US-09-569-098A-8
C 24	56.4	1.1	3095	6	5231168-1
25	56.2	1.1	665	2	US-08-883-795A-36
C 26	56.2	1.1	10640	4	US-09-417-485D-5
C 27	55.8	1.1	1241	1	US-07-593-657-6

C 28	55.8	1.1	1241	3	US-08-942-012B-3	Sequence 3, Appli
29	55.6	1.1	6265	4	US-09-129-112-3	Sequence 3, Appli
C 30	54.6	1.1	615	3	US-08-998-416-186	Sequence 186, App
C 31	54.6	1.1	636	3	US-08-998-416-1137	Sequence 1137, App
C 32	54.6	1.1	837	3	US-08-998-416-288	Sequence 288, App
33	54.4	1.1	6124	4	US-08-213-419B-3	Sequence 3, Appli
34	53.8	1.0	837	3	US-08-998-416-288	Sequence 288, App
35	53.8	1.0	912	3	US-08-688-988-3	Sequence 3, Appli
36	53.8	1.0	1511	1	US-07-991-867B-8	Sequence 8, Appli
37	53.8	1.0	1511	1	US-08-107-755A-8	Sequence 8, Appli
38	53.8	1.0	1511	2	US-08-544-332-8	Sequence 8, Appli
39	53.8	1.0	1511	4	US-09-370-861A-8	Sequence 8, Appli
C 40	53.8	1.0	2058	2	US-08-749-391-1	Sequence 1, Appli
C 41	53.8	1.0	2058	3	US-09-390-200-1	Sequence 1, Appli
42	53.6	1.0	636	3	US-08-998-416-1137	Sequence 1137, App
43	53.4	1.0	615	3	US-08-998-416-186	Sequence 186, App
44	53.4	1.0	660	1	US-07-991-867B-32	Sequence 32, Appl
45	53.4	1.0	660	1	US-08-107-755A-32	Sequence 32, Appl

ALIGNMENTS

RESULT 1  
US-09-313-294A-7428  
; Sequence 7428, Application US/09313294A  
; Patent No. 6476212  
; GENERAL INFORMATION:  
; APPLICANT: Lalgudi, Raghunath V.  
; APPLICANT: Ito, Laura Y.  
; APPLICANT: Sherman, Bradley K.  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
; FILE REFERENCE: PL-0017 US  
; CURRENT APPLICATION NUMBER: US/09/313,294A  
; CURRENT FILING DATE: 1999-05-14  
; NUMBER OF SEQ ID NOS: 7600  
; SOFTWARE: PERL Program  
; SEQ ID NO 7428  
; LENGTH: 299  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6476212 700381711H1  
; NAME/KEY: unsure  
; LOCATION: 42, 56  
; OTHER INFORMATION: a, t, c, g, or other  
US-09-313-294A-7428

Query Match 1.5%; Score 79.2; DB 4; Length 299;  
Best Local Similarity 70.9%; Pred. No. 1.5e-07;  
Matches 105; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 2919 AGGAGTAGAACTTCTGCTGATACCATGTGGAACCTCCGAACATATGTAGCTCCAGAGTAC 2978  
Db 85 AGGCTGATGGTCTGCTTCATACAAACATGTGGAACCCCAATATGTGCCCCCTGAGGTGA 144  
QY 2979 TTAGTGGACAGGGTTACGATGGTTTCAGCAGCTGATATTGGTCTTTCGGGGTTATCTTT 3038  
Db 145 TCGAGGATGGAGGCTATGATGGTGCACACTGCAGATATCTGGTCTTTCGGGAGTAATCTCT 204  
QY 3039 TCGTTATATTGGCTGGATATTACCTTT 3066  
Db 205 TTGTTCTTCTTGGTGGATATTACCTTT 232

RESULT 2  
US-08-487-826B-13  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan





RESULT 4  
US-09-056-075-1/c  
; Sequence 1, Application US/09056075  
; Patent No. 5955368  
; GENERAL INFORMATION:  
; APPLICANT: Johnson, Eric A.  
; APPLICANT: Bradshaw, Marite  
; APPLICANT: Rood, Julian  
; TITLE OF INVENTION: Expression System for Clostridium  
; TITLE OF INVENTION: Species  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Quarles & Brady  
; STREET: 1 South Pinckney Street  
; CITY: Madison  
; STATE: WI  
; COUNTRY: US  
; ZIP: 53701-2113  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/056,075  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J.  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.95238  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6243 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 3770..4013  
; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from  
; OTHER INFORMATION: plasmid RP4"  
US-09-056-075-1

Query Match 1.2%; Score 63.8; DB 2; Length 6243;  
Best Local Similarity 52.9%; Pred. No. 0.00042;  
Matches 137; Conservative 0; Mismatches 122; Indels 0; Gaps 0;  
Qy 523 ATCAAAATTAAGTATGATAACCTTTTAAACGGAATTGTTTGTAAAGTTTCATTACGTTA 582  
Db 1440 ATAAAAAATTTTTCAAACTTTAAATAAAAAAATATTTTATATATTTTATTTTTTA 1381  
Qy 583 TTTATATAAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAAT 642  
Db 1380 TTTTATATATTTTATATTTTATATTTTATATTTTATATTTTATATTTTATTTT 1321  
Qy 643 GCAAAATATTAAGTATGATAACCTTTTCAATCTTATCAATATTTTCTTATATTTCAATTA 702  
Db 1320 TATATTTTATTTTATATATTTTATATTTTATATTTTATATTTTATTTTATTTT 1261  
Qy 703 GCCACCAACCTACATTTTCAATTTCTCTTCAATTCATTAATTTCTGCTCTGTAT 762  
Db 1260 TTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1201  
Qy 763 CCTTCTTATCACTTCATT 781  
Db 1200 TTTTATTTTACCCCTCATT 1182

RESULT 5  
US-08-487-826B-13/c  
; Sequence 13, Application US/08487826B  
; Patent No. 5993827  
; GENERAL INFORMATION:  
; APPLICANT: Sim, Kim L.  
; APPLICANT: Chitnis, Chetan  
; APPLICANT: Miller, Louis H.  
; APPLICANT: Peterson, David S.  
; APPLICANT: Su, Xin-zhaun  
; APPLICANT: Wellem, Thomas E.  
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe Martens Olson & Bear  
; STREET: 620 Newport Center Drive 16th Floor  
; CITY: Newport Beach  
; STATE: California  
; COUNTRY: US  
; ZIP: 92660  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,826B  
; FILING DATE: 10-SEP-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Israel, Ned  
; REGISTRATION NUMBER: 29,655  
; REFERENCE/DOCKET NUMBER: NIH21.001CP1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 235-8550  
; TELEFAX: (619) 235-0176  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 19124 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-487-826B-13

Query Match 1.2%; Score 63.4; DB 2; Length 19124;  
Best Local Similarity 51.7%; Pred. No. 0.00065;  
Matches 168; Conservative 0; Mismatches 156; Indels 1; Gaps 1;  
Qy 521 TAATCAAAATTAAGTATGATAACCTTTTAAACGGAATTGTTTGTAAAGTTTCATTACGT 580  
Db 15915 TCATTAAATTTTAAATTTTATTTTAAATTTTATTTTAAATTTTATTTATTTATTTT 15856  
Qy 581 TATTTATATAAAAAACATAATGCAAAAGTACTAAAGTATAGTTATTTTATTTTATTTAATA 640  
Db 15855 TATTTAAATTTTATTTTATTTTATTTTATTTTATTTTAAATAAAATTTTATTTT 15796  
Qy 641 ATGCAAAATATTAAGTATGATAACCTTTTCAATCTTATCAATATTTTCTTATATTTCAAT 700  
Db 15795 ATGTATATATTTTATTTTAAACATTTTATTTTAAATTTTATTTTATTTTATGATATATTT 15736  
Qy 701 TAGCCACCAACCTACATTTTATTTCTCTTCAATTCATTTATTTTCTGCTCTTGT 760  
Db 15735 TATTTTAAATATATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 15676  
Qy 761 TTCCTTCTTATCATCTTCAATTTGCCAATTTTCTTCAATTTTCTTCAATTTGTCCTGTATG 820  
Db 15675 AGTTTTTTTTTCTTC-TTTTGTGTTTTTATTTTATAATCAATTTTATTTTATATA 15617

QY           821 TATGATTTTGAACATTAAATGTGA 845  
             | | | | |  
Db         15616 AAATTTTTTTTAATTTTTTTGA 15992

```

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZcgt-Fls
; PS-08-232-463-14

```

```

QY 2066 CTTTGCTAAGTTAAGTTTGCAGAGAACACAGACACTGGTGATAATGTAGCCATCAAAAT 2122
Db 1221 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1162
QY 2126 TATGGCTAAGACTACAATACTTAAGAACAGAAATGGTTGATCAGGTATGTTCTGGATTGTT 2185
Db 1161 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1102
QY 2186 TTTTACATGGAACTAAGTTGTTGCGTCAATGGTAT 2222
Db 1101 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1065

RESULT 7
US-07-867-106-2
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
US-07-867-106-2

```

QY 1223 TTTTGTGAACCTTGCTTGACTTAAATGAAGTTAAAAATTTGTAGTTAAAAATACAAAATTT



[illegible]

RESULT 10

```

US-09-129-112-3/c
; Sequence 3, Application US/09129112
; Patent No. 6465716
; GENERAL INFORMATION:
; APPLICANT: Etzler, Marilyn E.
; APPLICANT: Murphy, Judith B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
; FILE REFERENCE: 023070-079810US
; CURRENT APPLICATION NUMBER: US/09/129,112
; CURRENT FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: US 08/907,226
; PRIOR FILING DATE: 1997-08-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 6265
; TYPE: DNA
; ORGANISM: Dolichos biflorus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (633)..(944)
; NAME/KEY: intron
; LOCATION: (945)..(1022)
; NAME/KEY: exon
; LOCATION: (1023)..(1151)
; NAME/KEY: intron
; LOCATION: (1152)..(1559)
; NAME/KEY: exon
; LOCATION: (1560)..(1616)
; NAME/KEY: intron
; LOCATION: (1617)..(1697)
; NAME/KEY: exon
; LOCATION: (1698)..(1790)
US-09-129-112-3

```

Query Match	1.1%;	Score 57.2;	DB 4;	Length 6265;
Best Local Similarity	52.9%;	Pred. No. 0.0093;		
Matches 145;	Conservative	0;	Mismatches 128;	Indels 1; Gaps 1;
QY	1230	AACCTTGCTTGACCTAAATGAAGTTAAAAAATTTGTAGCTAAAAAATAGAAAAATTTTACTAA	1289	
Db	3709	AAATTTGCAGAAAAAATACATATGTATAATTAATAATATATATAATAACATAAATA	3650	
QY	1290	TTTGAGTTCGATTAAATCATACTCTAGATAAATTTGAAAAAATAATAAAATTTTGAAAA	1349	
Db	3649	TTAATATTTTATTATTATATATTATTAATAAAATAAAAATTAATAATAATTATTAA	3590	
QY	1350	TACTATATGTTTATTTTTTAAAAAATAATTTTACTAAATTCATAAGTCATATTAGATTAGTT	1409	
Db	3589	TAATATTTTATATTTTATATAATAATAAAATTTAAAAAATAATAATAAAATTTTAAAA	3530	
QY	1410	TTTTTTTCTTTTTTAAATTTTGAAAAACCTCACAAATTATTAAATTTGAAAGATTCCAATG	1469	
Db	3529	ATAAATATTTTATTTTTCATAAT-AATAAATAAATTTAAAAATAATTTAAAAATAATAA	3471	
QY	1470	CAATGTTAGTTTTTTAAAAAGTTTAAATCACCCAAAC	1503	
Db	3470	TTATTCAACTTTTTTCTCTTTCTTAATAATATCCAAAC	3437	

RESULT 11

US-07-867-106-2/c

```

; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Slade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526-ris
; STREET: One Liberty Place 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/867,106
; FILING DATE: 19920625
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PJ 7187
; APPLICATION NUMBER: PCT/AU90/00530
; FILING DATE: 02-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Feeney, Joanne Longo
; REGISTRATION NUMBER: 35,134
; REFERENCE/DOCKET NUMBER: RICE-0002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5852 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2378..5038
; NAME/KEY: CDS
; LOCATION: 2378..5038
; US-07-867-106-2

Query Match          1.1%; Score 57; DB 1; Length 5852;
Best Local Similarity 52.3%; Pred. No. 0.01;
Matches 150; Conservative 0; Mismatches 135; Indels 2; Gaps 1;

QY 1158 TTTGTAGCATTCGAAAGTATCTTTTGTATGCTTTTGCTTAAACCTTAGACATCCTTTGTTCTC 1217
Db 515 TTTTGGGCATTTTTTTTTCACCTTTTTTTTTCATTTTCCACCTTTTTTTCACCTTTTTTTTTCCT 456

QY 1218 TTGGTTTTTTTGAACCTTGCTTTGACTTAAATGAAGTTAAAAATTTGTAGTTAAAAATAGA 1277
Db 455 TTTTTTCCTTTTTTTTTTTTTTTAAAGTTAAATAAATAAATAAATAAATAAATAAATAAATAA 396

QY 1278 AAATTTTACTAATTTGAGTTCGATTAATCATAGTCTAGATAAATTTGAAAAAAATTTAAAT 1337
Db 395 TTATTTATATATCTTCATGACCTTCATT--TTTTTAAATTTTATGTAGTTAATTTATAA 338

QY 1338 AAATTTTGAAAAATACTATATGTTTATTTTTTAAAAATAAATTTACTAAATTTGATAAGTGAT 1397
Db 337 AAATTGTGATCTTTTAAATCTATACTATTTTTTAAATAATTTTAAAAAGTTATTAAAAACAAGA 278

QY 1398 ATTAGATTAGTTTTTTTTTTTCTTTTTTTTAAATTTTTTGAACCTTCACAA 1444

```





```
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTTT
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 5903
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6019
; OTHER INFORMATION: 10-346-141 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6141
; OTHER INFORMATION: 10-346-263 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6183
; OTHER INFORMATION: 10-346-305 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6338
; OTHER INFORMATION: 10-347-74 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6375
; OTHER INFORMATION: 10-347-111 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6429
; OTHER INFORMATION: 10-347-165 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6467
; OTHER INFORMATION: 10-347-203 : polymorphic base A or G
; NAME/KEY: allele
```

```
; LOCATION: 6484
; OTHER INFORMATION: 10-347-220 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6534
; OTHER INFORMATION: 10-347-271 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 6611
; OTHER INFORMATION: 10-347-348 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 7668
; OTHER INFORMATION: 10-348-391 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8608
; OTHER INFORMATION: 10-349-47 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 8658
; OTHER INFORMATION: 10-349-97 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8703
; OTHER INFORMATION: 10-349-142 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 8777
; OTHER INFORMATION: 10-349-216 : deletion of CTG
; NAME/KEY: allele
; LOCATION: 8785
; OTHER INFORMATION: 10-349-224 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 8926
; OTHER INFORMATION: 10-349-368 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12171
; OTHER INFORMATION: 10-350-72 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12429
; OTHER INFORMATION: 10-350-332 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13341
; OTHER INFORMATION: 10-507-170 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 13492
; OTHER INFORMATION: 10-507-321 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 13524
; OTHER INFORMATION: 10-507-353 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13535
```

Query Match 1.1%; Score 57; DB 4; Length 20674;  
Best Local Similarity 49.0%; Pred. No. 0.013;  
Matches 179; Conservative 0; Mismatches 185; Indels 1; Gaps 1;

```
QY 1129 AATCCTTATCGAACCTTTTATATATATCCTTTGTAGCATTTGGAAAGTATCTTTGTATGCT 1188
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11171 AATAAATAATTAAATTTAAATAAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 11230

QY 1189 TTGTCTTAAACCTTAGACATCCTTGTCTCTCTCTTGGTTTTTTTGAACCTTGCCTTGACTTAAAT 1248
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11231 CATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 11290

QY 1249 GAAGTTAAAAATTTGTAGTTAAAAATAGAAAAATTTTACTAAATTTGAGITTCGATTAAATCAT 1308
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11291 TAATTTAATATTAAATTTAACTTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 11350

QY 1309 AGCTAGATAAATTTGAAAAAAATTTAAATAAATTTTGAAAAATACTATATGTTTTTTT 1368
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11351 TTTAATTTTAAATATTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 11410

QY 1369 AAAAATAATTTTACTAAATTTGATAAGTGATATAGATTAGTTTTTTTTTTTTTTTTTCTTTTAAAT 1428
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11411 TTAATTGCTTCAATATTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAAATTT 11470

QY 1429 TT-TGAAAAACCTCACAAATTATTAAATTTGAAAGATTCAAATGCAATGTTAGTTTTTTTAAAA 1487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```



```
;
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1128
; OTHER INFORMATION: 10-508-191 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1182
; OTHER INFORMATION: 10-508-245 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTTTT
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 5903
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6019
; OTHER INFORMATION: 10-346-141 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6141
; OTHER INFORMATION: 10-346-263 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6183
; OTHER INFORMATION: 10-346-305 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6338
; OTHER INFORMATION: 10-347-74 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6375
; OTHER INFORMATION: 10-347-111 : polymorphic base G or C
```

```
;
; NAME/KEY: allele
; LOCATION: 6429
; OTHER INFORMATION: 10-347-165 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6467
; OTHER INFORMATION: 10-347-203 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6484
; OTHER INFORMATION: 10-347-220 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6534
; OTHER INFORMATION: 10-347-271 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 6611
; OTHER INFORMATION: 10-347-348 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 7668
; OTHER INFORMATION: 10-348-391 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8608
; OTHER INFORMATION: 10-349-47 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 8658
; OTHER INFORMATION: 10-349-97 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 8703
; OTHER INFORMATION: 10-349-142 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 8777
; OTHER INFORMATION: 10-349-216 : deletion of CTG
; NAME/KEY: allele
; LOCATION: 8785
; OTHER INFORMATION: 10-349-224 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 8926
; OTHER INFORMATION: 10-349-368 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12171
; OTHER INFORMATION: 10-350-72 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12429
; OTHER INFORMATION: 10-350-332 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13341
; OTHER INFORMATION: 10-507-170 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 13492
; OTHER INFORMATION: 10-507-321 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 13524
; OTHER INFORMATION: 10-507-353 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 13535
```

Query Match 1.1%; Score 56.8; DB 4; Length 20674;  
Best Local Similarity 49.7%; Pred. No. 0.015;  
Matches 145; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

```
QY 1245 AAATGAAGTTAAAAATTTGTAGTTAAAAATAGAAAAATTTTACTAATTTGAGTTCCGATTAA 1304
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11349 ATATTAAAAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11290

QY 1305 TCATAGTCTAGATAATTTGAAAAAAAATTAATAATTTTGAAAAATACATATGTTTATT 1364
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11289 ATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGT 11230

QY 1365 TTTTAAAAAATAATTTACTAAATTTGATAAGTGATATTAGATTAGTTTTTTTTTTTCTTTT 1424
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11229 TTATTTAAAAATTAATTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTATTTT 11170

QY 1425 AAATTTTGAAAACCTCACAAATTATTAAATGAAAGATTCAAAATGCAATGCTAGTTTTTA 1484
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11169 AAAATTAAAAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11110
```



Qy 1485 AAAGTTTAATCACCCAAACGGATAATTGACCCGAACGACTAATTCAGGTCGT 1536

Dd 11109 TTATTAAATTAAAAATATTAAATAATTTCTTAGCTATTAAAAAAGGCAT 11058

Search completed: October 21, 2003, 09:34:02  
Job time : 269 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 21, 2003, 00:09:07 ; Search time 1217 seconds  
(without alignments)  
11409.950 Million cell updates/sec

Title: US-09-824-735-1  
Perfect score: 5144  
Sequence: 1 tttttgcggaataatctcg.....acgttttgatcccaacttaa 5144

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	5144	100.0	5144	24	Gene encoding Arab
2	414.6	8.1	2000	24	Arabidopsis thalia
3	401.4	7.8	2000	24	Arabidopsis thalia
4	371.2	7.2	2000	24	Arabidopsis thalia
5	234	4.5	1082138	21	Arabidopsis thalia
6	206.8	4.0	1377	24	cDNA encoding Arab
7	183.2	3.6	1341	21	Arabidopsis thalia
8	163.6	3.2	163319	21	Arabidopsis thalia

C	9	152.4	3.0	2000	24	ABZ17294	Arabidopsis thalia
C	10	115.8	2.3	1502	24	AAL48584	A thaliana WRKY64
C	11	112.4	2.2	1082138	21	AAF22305	Arabidopsis thalia
	12	105	2.0	8056	25	ABZ10246	Haematopoietic cel
	13	97.6	1.9	404	25	ABX21874	Human GDP-mannose
	14	97.2	1.9	397	25	ABX62736	Arabidopsis thalia
	15	97.2	1.9	1338	21	AAC42994	Arabidopsis thalia
	16	97.2	1.9	1590	21	AAC36536	Arabidopsis thalia
	17	93.8	1.8	8056	25	ABZ10100	Haematopoietic cel
	18	92.2	1.8	364	21	AAA31238	Plant microsateili
	19	91.2	1.8	845	21	AAC46643	Zea mays DNA fragm
C	20	88.4	1.7	8056	25	ABZ10246	Haematopoietic cel
	21	88	1.7	742	24	ABQ65474	Arabidopsis thalia
	22	84	1.6	1869	21	AAAS2775	Soybean putative c
	23	84	1.6	1869	25	ACC00752	Glycine max oil tr
	24	83.8	1.6	1707	21	AAA52767	Corn putative cata
	25	83.8	1.6	1707	25	ACC00742	Zea mays oil trait
C	26	83.8	1.6	8056	25	ABZ10100	Haematopoietic cel
	27	83.6	1.6	465	21	AAC38241	Zea mays DNA fragm
	28	83.2	1.6	1073	21	AAC51017	Arabidopsis thalia
	29	83.2	1.6	1622	21	AAC44238	Arabidopsis thalia
C	30	83.2	1.6	103929	21	AAF22287	BAC containing rep
	31	82.2	1.6	1326	24	ABZ12267	Arabidopsis thalia
	32	81.8	1.6	15548	24	ABL34155	Human immune syste
	33	81.2	1.6	7814	22	AAS46530	Arabidopsis thalia
	34	80.4	1.6	1320	21	AAC46102	Human immune syste
	35	80.4	1.6	15387	24	ABL32184	Corn ear-derived p
	36	79.2	1.5	299	25	ABX88968	Human chemically m
	37	78.4	1.5	8067	24	ABN80225	Corn putative cata
	38	77.2	1.5	512	21	AAAS2778	Human immune syste
	39	76.6	1.5	7306	24	ABL33637	Zea mays DNA fragm
	40	75.6	1.5	469	21	AAC43576	Arabidopsis thalia
	41	75.2	1.5	1326	24	ABZ14508	Arabidopsis thalia
	42	74.8	1.5	1198	21	AAC49879	Arabidopsis thalia
	43	74.6	1.5	1470	24	ABZ14173	Arabidopsis thalia
	44	74.2	1.4	5893	24	ABL32859	Human immune syste
	45	74	1.4	83391	24	ABQ67094	Human angiogenesis

ALIGNMENTS

RESULT 1  
ID ABK91073 standard; DNA; 5144 BP.

AC ABK91073;

DT 15-NOV-2002 (first entry)

DE Gene encoding Arabidopsis thaliana SOS2 serine/threonine protein kinase.

KW Salt overly sensitive 7; SOS2; serine/threonine protein kinase;

KW salt tolerance; agricultural crop; rice; corn; wheat; cotton;

KW peanut; soybean; plant protectant; plant; transgenic; gene; ds.

OS Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 2004..5133

FT /tag= a

FT /product= "SOS2"

FT exon 2004..2168

FT /tag= b

FT /number= 1

FT intron 2169..2254

FT /tag= c

FT /number= 1

FT exon 2255..2317

FT /tag= d

FT /number= 2

FT intron 2318..2415

FT /tag= e

FT /number= 2  
FT 2416..2487  
FT /\*tag= f  
FT /number= 3  
FT 2488..2572  
FT /\*tag= g  
FT /number= 3  
FT 2573..2680  
FT /\*tag= h  
FT /number= 4  
FT 2681..2777  
FT /\*tag= i  
FT /number= 4  
FT 2778..2852  
FT /\*tag= j  
FT /number= 5  
FT 2853..2920  
FT /\*tag= k  
FT /number= 5  
FT 2921..3100  
FT /\*tag= l  
FT /number= 6  
FT 3101..3224  
FT /\*tag= m  
FT /number= 6  
FT 3225..3314  
FT /\*tag= n  
FT /number= 7  
FT 3315..3417  
FT /\*tag= o  
FT /number= 7  
FT 3418..3540  
FT /\*tag= p  
FT /number= 8  
FT 3541..4051  
FT /\*tag= q  
FT /number= 8  
FT 4052..4171  
FT /\*tag= r  
FT /number= 9  
FT 4172..4399  
FT /\*tag= s  
FT /number= 9  
FT 4400..4516  
FT /\*tag= t  
FT /number= 10  
FT 4517..4734  
FT /\*tag= u  
FT /number= 10  
FT 4735..4791  
FT /\*tag= v  
FT /number= 11  
FT 4792..4878  
FT /\*tag= w  
FT /number= 11  
FT 4879..4953  
FT /\*tag= x  
FT /number= 12  
FT 4954..5037  
FT /\*tag= y  
FT /number= 12  
FT 5038..5130  
FT /\*tag= z  
FT /number= 13  
XX US2002095032-A1.  
PN 18-JUL-2002.  
XX 04-APR-2001; 2001US-0824735.  
PF 04-APR-2000; 2000US-194649P.  
XX

PA (UYAR-) UNIV ARIZONA.  
XX  
PI Zhu J, Liu J, Ishitani M, Kim C, Halfter U;  
XX  
DR WPI; 2002-665801/71.  
XX P-PSDB; ABG31356.  
PT New SOS2 protein kinase polypeptide and nucleic acids encoding the  
PT polypeptide, useful for increasing salt tolerance to plants, especially  
PT to agricultural crops -  
XX  
PS Claim 3; Page 9-12; 28pp; English.  
XX  
CC The present invention relates to the isolation of an Arabidopsis  
CC thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine  
CC protein kinase. The polynucleotide sequence encoding SOS2 can be used  
CC to increase salt tolerance in plants, particularly agricultural crops  
CC such as rice, corn, wheat, cotton, peanut and soybean. The  
CC polynucleotide sequence encoding SOS2 can also be used to generate  
CC transgenic plants, as hybridization probes for RNA, cDNA and DNA to  
CC isolate those cDNAs or genes which exhibit a high degree of similarity  
CC to the sequence of the SOS2 gene, and as PCR primers for the production  
CC of DNA which encodes an enzyme having serine/threonine protein kinase  
CC activity. The present sequence encodes Arabidopsis thaliana SOS2  
CC serine/threonine protein kinase.  
XX  
SQ Sequence 5144 BP; 1523 A; 851 C; 943 G; 1827 T; 0 other;  
  
Query Match 100.0%; Score 5144; DB 24; Length 5144;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 5144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TTTTGGCGGAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGTTTACGTTTT 60  
Db |||||  
1 TTTTGGCGGAAATCTCGGTTTACGTTTTTGGCGGAAAAATCTCGTTTACGTTTT 60  
  
Qy 61 TGGCGGAAAAATCTCGGTTTACGTTTTTGGGAGAAAAATCACGGTTTACTTTTTTTG 120  
Db |||||  
61 TGGCGGAAAAATCTCGGTTTACGTTTTTGGGAGAAAAATCACGGTTTACTTTTTTTG 120  
  
Qy 121 CGGAAAAATCACGGATTATGTTTTTTGGTGAAAAAATTACGAGTTTACTTTTCTCAAT 180  
Db |||||  
121 CGGAAAAATCACGGATTATGTTTTTTGGTGAAAAAATTACGAGTTTACTTTTCTCAAT 180  
  
Qy 181 TTCATTGCTTGATATTTAAGAAATTTGGAAAAATTAATTTTATTAAATGGTTTAGA 240  
Db |||||  
181 TTCATTGCTTGATATTTAAGAAATTTGGAAAAATTAATTTTATTAAATGGTTTAGA 240  
  
Qy 241 TGTGTTGGTTAAACCTAAATGGCATTTGGTTTAGAGATTTTAGTTGGTTTATTCAATT 300  
Db |||||  
241 TGTGTTGGTTAAACCTAAATGGCATTTGGTTTAGAGATTTTAGTTGGTTTATTCAATT 300  
  
Qy 301 TACAAAAATTGATGGTTAATTTGGATAAACCATGGAACCATTAACCATTAACAACCTAAC 360  
Db |||||  
301 TACAAAAATTGATGGTTAATTTGGATAAACCATGGAACCATTAACCATTAACAACCTAAC 360  
  
Qy 361 TCATTTTACTCATCAACCAATTGACTCATCACTCATCTCACTCACTCACTCACTTTGAG 420  
Db |||||  
361 TCATTTTACTCATCAACCAATTGACTCATCACTCATCTCACTCACTCACTCACTTTGAG 420  
  
Qy 421 TCAAAAAATTTCAACTATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480  
Db |||||  
421 TCAAAAAATTTCAACTATTAGGGTTTCATGGGTTGAGTTGAGTTGAGTTGAGTTGAGTT 480  
  
Qy 481 TTGACCCCATTTTGACACCCCTACATATGATCAATAAGTTAATAATCAAAAATTACTATTGA 540  
Db |||||  
481 TTGACCCCATTTTGACACCCCTACATATGATCAATAAGTTAATAATCAAAAATTACTATTGA 540  
  
Qy 541 TAACTTTTTAACGGAATTTTGTAAAGTTTCACTTACGTTATTTATATAAAAAACATA 600  
Db |||||  
541 TAACTTTTTAACGGAATTTTGTAAAGTTTCACTTACGTTATTTATATAAAAAACATA 600  
  
Qy 601 ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATGCAAAAATATTACTGTAA 660

Db 601 |||||ATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGCAAAAATATTAAGTAA 660  
Qy 661 TACTTTTTCATTCTTATCAATATTTTCTTATATTTCAATAGCACCAACACCTACAT 720  
Db 661 |||||TACTTTTTCATTCTTATCAATATTTTCTTATATTTCAATAGCACCAACACCTACAT 720  
Qy 721 TCATATTTCTCTTACATTCATTTATTTCTTGCTCTTGATTCCTTATCAATCTTCA 780  
Db 721 TCATATTTCTCTTACATTCATTTATTTCTTGCTCTTGATTCCTTATCAATCTTCA 780  
Qy 781 TGCCAAATTTTCAATGACATTTGTCATGCTTACTTTGTATGATGATTTTGAACATTTAA 840  
Db 781 TGCCAAATTTTCAATGACATTTGTCATGCTTACTTTGTATGATGATTTTGAACATTTAA 840  
Qy 841 TGTGAATCCCAACCGACTATAGGACTATCAAGAAAGTTTTCAAAACCTTTTAAAAAAGATCT 900  
Db 841 TGTGAATCCCAACCGACTATAGGACTATCAAGAAAGTTTTCAAAACCTTTTAAAAAAGATCT 900  
Qy 901 TGAACCTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCCCTCTTCTTGTGATGA 960  
Db 901 TGAACCTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCCCTCTTCTTGTGATGA 960  
Qy 961 AGCTCTTCTCGTCGGCGGCAAAATCTAGGCGATATAAGCCCTCTTCAACATCACCTAGTAT 1020  
Db 961 AGCTCTTCTCGTCGGCGGCAAAATCTAGGCGATATAAGCCCTCTTCAACATCACCTAGTAT 1020  
Qy 1021 ATTGACCGTGACCATCTTTTGGACCATTTGTCATGTAATGAACCGTCGATAAACCGTGTT 1080  
Db 1021 ATTGACCGTGACCATCTTTTGGACCATTTGTCATGTAATGAACCGTCGATAAACCGTGTT 1080  
Qy 1081 ATCACTTACGCCCAAAATTTTCCCTAGTCTTTTCCAAACTCTCACGAATCCTTATCGA 1140  
Db 1081 ATCACTTACGCCCAAAATTTTCCCTAGTCTTTTCCAAACTCTCACGAATCCTTATCGA 1140  
Qy 1141 ACTTTTATATATACATTTGTAGCATTTGGAAGTATCTTTGATGCTTTGCTTAAACT 1200  
Db 1141 ACTTTTATATATACATTTGTAGCATTTGGAAGTATCTTTGATGCTTTGCTTAAACT 1200  
Qy 1201 TAGACATCCTTGTCTTGGTTTGTGAACTTGTGACTTAAATGAAGTTAAAAAT 1260  
Db 1201 TAGACATCCTTGTCTTGGTTTGTGAACTTGTGACTTAAATGAAGTTAAAAAT 1260  
Qy 1261 TTGTAGTTAAATAGAAAATTTTACTAATTTGAGTTTGAATTAATCATAGTCTAGATAAT 1320  
Db 1261 TTGTAGTTAAATAGAAAATTTTACTAATTTGAGTTTGAATTAATCATAGTCTAGATAAT 1320  
Qy 1321 TTGAAAAAAATTAATAATAAATTTTGAAAAATCTATATGTTTATTTTAAAAAATAATTA 1380  
Db 1321 TTGAAAAAAATTAATAATAAATTTTGAAAAATCTATATGTTTATTTTAAAAAATAATTA 1380  
Qy 1381 CTAAATTTGATAAGTGATATAGATTTAGTTTCTTTTCTTTTAAATTTTGAAAAACCTC 1440  
Db 1381 CTAAATTTGATAAGTGATATAGATTTAGTTTCTTTTCTTTTAAATTTTGAAAAACCTC 1440  
Qy 1441 ACAATTTAATAATGAAAAGATTCAAAATGCAATGTTAGTTTAAAAAGTTTAAATCACCCA 1500  
Db 1441 ACAATTTAATAATGAAAAGATTCAAAATGCAATGTTAGTTTAAAAAGTTTAAATCACCCA 1500  
Qy 1501 AACGGATAATTGACCGAAGACTAATTCAGTTCGTATACGGGTACAGTCAAAATAACCCG 1560  
Db 1501 AACGGATAATTGACCGAAGACTAATTCAGTTCGTATACGGGTACAGTCAAAATAACCCG 1560  
Qy 1561 ACCCGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTCACGGCTTTCGCGA 1620  
Db 1561 ACCCGAAATGCTCAACGGATCTGCACGGACGTTTCGCATCGACGATTCACGGCTTTCGCGA 1620  
Qy 1621 ATCGCATCACGAGCCTTCTCTCTCTACGCTCTTTTCTCAACACCTTCTCTGCGAATC 1680  
Db 1621 ATCGCATCACGAGCCTTCTCTCTCTACGCTCTTTTCTCAACACCTTCTCTGCGAATC 1680  
Qy 1681 CAATTTGGTATTTTACGATTTCTTTTCCGACTAGCCCAATTCGCAATTCAGCCGTACGA 1740  
|||

Db 1681 CAATTTGGTATTTTACGATTTCTCTTTCGACTACGCCCAATTCGCAATTCAGCCGTACGA 1740  
Qy 1741 ATTTTATATATATGATATGATCTTGATCTTACCTTTTTCGTCGTTTCTTGATCGAATTT 1800  
Db 1741 ATTTTATATATATGATATGATCTTGATCTTACCTTTTTCGTCGTTTCTTGATCGAATTT 1800  
Qy 1801 ATGTTGATTTATGCTCGCAATCAAAATTTATGATACAGAGTAATTTTGTATATATGAT 1860  
Db 1801 ATGTTGATTTATGCTCGCAATCAAAATTTATGATACAGAGTAATTTTGTATATATGAT 1860  
Qy 1861 GTGTAGGAGCTAAAAATTAGCAAAACAATACGTAAAGTAATTTGAAATCGAAATCATAAACGT 1920  
Db 1861 GTGTAGGAGCTAAAAATTAGCAAAACAATACGTAAAGTAATTTGAAATCGAAATCATAAACGT 1920  
Qy 1921 TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTTGATAGCTTGTGGTCAATG 1980  
Db 1921 TTAAGGAAAGAGGTTTTTACTAAGTCTCTGAATAATCTGATTTGATAGCTTGTGGTCAATG 1980  
Qy 1981 GATCAGATAAAAGTTTTGTAAAGAAATGACAAAAGAAAATGAGAAAGAGTGGGCAAGTACGAGG 2040  
Db 1981 GATCAGATAAAAGTTTTGTAAAGAAATGACAAAAGAAAATGAGAAAGAGTGGGCAAGTACGAGG 2040  
Qy 2041 TTGGTCGCACATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTTCGAGAGAACACAGACA 2100  
Db 2041 TTGGTCGCACATAGGTGAAGGAACCTTTGCTAAGGTTAAGTTTTCGAGAGAACACAGACA 2100  
Qy 2101 CTGGTGATAATGTAGCCCATCAAAATTTATGGCTAAGAGTACATACTTAAGAACAGAAATGG 2160  
Db 2101 CTGGTGATAATGTAGCCCATCAAAATTTATGGCTAAGAGTACATACTTAAGAACAGAAATGG 2160  
Qy 2161 TTGATCAGGTATGTTCTGGAATGTTTTTACATGGAACCTAAGGTTGTTGCGTCAATGGT 2220  
Db 2161 TTGATCAGGTATGTTCTGGAATGTTTTTACATGGAACCTAAGGTTGTTGCGTCAATGGT 2220  
Qy 2221 ATGATCTTTGATTTCTGTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA 2280  
Db 2221 ATGATCTTTGATTTCTGTTTAAAGCTCTTTTACAGATAAAAGAGAGATATCTATAATGAA 2280  
Qy 2281 GATTTGTTCTGTCACCGGAACATAGTGAGGTTGATGAGGTATGTTGTTTCCATGCA 2340  
Db 2281 GATTTGTTCTGTCACCGGAACATAGTGAGGTTGATGAGGTATGTTGTTTCCATGCA 2340  
Qy 2341 TCTGCGAAATTTTATCTCTGAAAGTGTGTTTGCATCATTTGTTCTCTGTTGTTTTTTG 2400  
Db 2341 TCTGCGAAATTTTATCTCTGAAAGTGTGTTTGCATCATTTGTTCTCTGTTGTTTTTTG 2400  
Qy 2401 ATTTTCCCGATGATGAGTGTGCGGAGTCTTCGAAAATATATATAGTTTTGGAGTTTGG 2460  
Db 2401 ATTTTCCCGATGATGAGTGTGCGGAGTCTTCGAAAATATATATAGTTTTGGAGTTTGG 2460  
Qy 2461 ACAGGAGGAGGCTCTTTTGATAGATTTGTACGGAACTTCCATACCTTTGTAGGCAGCTCCA 2520  
Db 2461 ACAGGAGGAGGCTCTTTTGATAGATTTGTACGGAACTTCCATACCTTTGTAGGCAGCTCCA 2520  
Qy 2521 TTAGTTAAACCTCTCTACTTAAATTTTAAATATATATGAAATCTTTTCATGCAGGTTCA 2580  
Db 2521 TTAGTTAAACCTCTCTACTTAAATTTTAAATATATATGAAATCTTTTCATGCAGGTTCA 2580  
Qy 2581 AGGGAGGCTTGAAGAAAGTGAGTCTCGGAAATCTTTTCAACAGCTTTGTAGATGCTGTGC 2640  
Db 2581 AGGGAGGCTTGAAGAAAGTGAGTCTCGGAAATCTTTTCAACAGCTTTGTAGATGCTGTGC 2640  
Qy 2641 TCATTGTCACTGCAAGGTTGTTTACACCGTGACCTAAAGTAAAGACGTTGTTTTGTTTT 2700  
Db 2641 TCATTGTCACTGCAAGGTTGTTTACACCGTGACCTAAAGTAAAGACGTTGTTTTGTTTT 2700  
Qy 2701 ACCAATATTCCTCAGAAATATCTCACTGCTTGAATCCAGACTTGATATTTTGTGTCGC 2760  
Db 2701 ACCAATATTCCTCAGAAATATCTCACTGCTTGAATCCAGACTTGATATTTTGTGTCGC 2760  
Qy 2761 TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATGGAATCTGAAGGTTT 2820  
Db 2761 TATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAATGGAATCTGAAGGTTT 2820



QY	2821	CGGATTTTCGGA	CTCAGTGCAT	TGCCTCAGGAAGT	AAGTGCTCTTATCTCTGCTTCAGCAG	2880
Db	2821	CGGATTTTCGGA	CTCAGTGCAT	TGCCTCAGGAAGT	AAGTGCTCTTATCTCTGCTTCAGCAG	2880
QY	2881	TCTGCTTACGTGGT	CATTAACT	TGTGTATATACTCAAT	CAGGAGTAGAACTTCTGCGTAC	2940
Db	2881	TCTGCTTACGTGGT	CATTAACT	TGTGTATATACTCAAT	CAGGAGTAGAACTTCTGCGTAC	2940
QY	2941	CACATGTGGA	ACTCCGAAC	TATAGTCTCCAGAGGTACT	TAGTGGACAGGGTTACGATGG	3000
Db	2941	CACATGTGGA	ACTCCGAAC	TATAGTCTCCAGAGGTACT	TAGTGGACAGGGTTACGATGG	3000
QY	3001	TTCAGCAGCTGAT	ATTGGTCT	TTCGGGGTTATTCT	TTTTCGTTATATTGGCTGGATATTT	3060
Db	3001	TTCAGCAGCTGAT	ATTGGTCT	TTCGGGGTTATTCT	TTTTCGTTATATTGGCTGGATATTT	3060
QY	3061	ACCTTTTTCG	GACGGATCT	CCAGGGTTGTACAGAAAAGT	AAAGTAAACATATCTTTTCGG	3120
Db	3061	ACCTTTTTCG	GACGGATCT	CCAGGGTTGTACAGAAAAGT	AAAGTAAACATATCTTTTCGG	3120
QY	3121	GAAGAAATCAT	GAATTCCT	TGTGCATGGCTTTTGT	CAAACCGTTTATTGATTTGGTTTTGC	3180
Db	3121	GAAGAAATCAT	GAATTCCT	TGTGCATGGCTTTTGT	CAAACCGTTTATTGATTTGGTTTTGC	3180
QY	3181	AAATTCACCC	TTAGTTTT	TGAGCTTTTACACAT	TGTTATTACAGATAAAATGCAGCAGAT	3240
Db	3181	AAATTCACCC	TTAGTTTT	TGAGCTTTTACACAT	TGTTATTACAGATAAAATGCAGCAGAT	3240
QY	3241	TTTCTGTCC	ACCGTGGTTT	CCGAGAAAGTGAAGTTT	TTTAATACATAGGATCTTGACC	3300
Db	3241	TTTCTGTCC	ACCGTGGTTT	CCGAGAAAGTGAAGTTT	TTTAATACATAGGATCTTGACC	3300
QY	3301	CCAATCCC	AAAACAGTGAGT	ATTGCTTTGTTCTCTCC	TAGCTATCAGGTTTGGTGAT	3360
Db	3301	CCAATCCC	AAAACAGTGAGT	ATTGCTTTGTTCTCTCC	TAGCTATCAGGTTTGGTGAT	3360
QY	3361	ATTTAATGTT	CTAGTAATT	TATATCTGTTTATCT	ATTATTGTTTCTCAATTAGACGCGT	3420
Db	3361	ATTTAATGTT	CTAGTAATT	TATATCTGTTTATCT	ATTATTGTTTCTCAATTAGACGCGT	3420
QY	3421	ATTCAAAT	TTCAAGGAAT	CAAGAAAGATCCT	TGGTTCAGATTAATATGTGCCATACGA	3480
Db	3421	ATTCAAAT	TTCAAGGAAT	CAAGAAAGATCCT	TGGTTCAGATTAATATGTGCCATACGA	3480
QY	3481	GCAAGGG	AAGAAGTGAAT	TTGGATGATATT	CGTGCAGTTTGTGATGGAATTGAG	3540
Db	3481	GCAAGGG	AAGAAGTGAAT	TTGGATGATATT	CGTGCAGTTTGTGATGGAATTGAG	3540
QY	3541	GTTTGTGCT	TCGCCTTCAT	TATATGCTCTTT	TGCTGGTCAATTCCTTTTAAATGTTAAGAT	3600
Db	3541	GTTTGTGCT	TCGCCTTCAT	TATATGCTCTTT	TGCTGGTCAATTCCTTTTAAATGTTAAGAT	3600
QY	3601	CTCTTAGG	AACGTTTGGATG	ACCAGAAAGTGT	TGCTACAGGATAGAACAAAATAGT	3660
Db	3601	CTCTTAGG	AACGTTTGGATG	ACCAGAAAGTGT	TGCTACAGGATAGAACAAAATAGT	3660
QY	3661	AGGCATG	TGTTAGTTAC	CAAACCTGTAACT	GCTTCTTATTCAATTCCGCAACCATAG	3720
Db	3661	AGGCATG	TGTTAGTTAC	CAAACCTGTAACT	GCTTCTTATTCAATTCCGCAACCATAG	3720
QY	3721	ACCTTAGG	AAGACTTAGATT	TACAAGAGAATT	CTCTATTTCTCGACCAAAACCCCTAGACA	3780
Db	3721	ACCTTAGG	AAGACTTAGATT	TACAAGAGAATT	CTCTATTTCTCGACCAAAACCCCTAGACA	3780
QY	3781	AAATCCAG	AATACCCCTAG	CTAATTACAAT	GTTCATGTACCTATCAATATATATCTCC	3840
Db	3781	AAATCCAG	AATACCCCTAG	CTAATTACAAT	GTTCATGTACCTATCAATATATATCTCC	3840
QY	3841	TGCTAAG	ATTGCTTAGCT	TTGGTATAGCCT	TAGATATATAGATACCGTAAATTTCTAAT	3900
Db	3841	TGCTAAG	ATTGCTTAGCT	TTGGTATAGCCT	TAGATATATAGATACCGTAAATTTCTAAT	3900

QY	3901	GCATATTT	TAGAAATGTTT	TACTTAAACT	CAGGCTCTTGCTCTTCTTAA	ACTTGTTCAA	3960
Db	3901	GCATATTT	TAGAAATGTTT	TACTTAAACT	CAGGCTCTTGCTCTTCTTAA	ACTTGTTCAA	3960
QY	3961	TTGTTAA	AACTAAACCT	CAGTATCTGTCT	TAGCTAAAGTTACTTTT	TACTTTTCATT	4020
Db	3961	TTGTTAA	AACTAAACCT	CAGTATCTGTCT	TAGCTAAAGTTACTTTT	TACTTTTCATT	4020
QY	4021	AAGTTG	ACCTGTCAAT	TGCAATTT	CACAGGCGAGTTAT	GTAGAGAG	4080
Db	4021	AAGTTG	ACCTGTGTCAAT	TGCAATTT	CACAGGCGAGTTAT	GTAGAGAG	4080
QY	4081	AAATGAT	GAAGGGCCCT	CGATGAATGCCT	TTTGAGATGATTACCT	TATCACAAGGCTT	4140
Db	4081	AAATGAT	GAAGGGCCCT	CGATGAATGCCT	TTTGAGATGATTACCT	TATCACAAGGCTT	4140
QY	4141	AAATTTAT	CTGCAC	TATTTGACAGGCGAC	CAGGTAGTACTGATTTT	CTATTACTGGTCA	4200
Db	4141	AAATTTAT	CTGCAC	TATTTGACAGGCGAC	CAGGTAGTACTGATTTT	CTATTACTGGTCA	4200
QY	4201	AGAGAT	CTCCATTT	TCGAATAAAGAA	TGTCGGTAGCATCTAT	CTTCAGACTGCCCGTTT	4260
Db	4201	AGAGAT	CTCCATTT	TCGAATAAAGAA	TGTCGGTAGCATCTAT	CTTCAGACTGCCCGTTT	4260
QY	4261	TGACTG	CCCTTATGAT	GCCTGTCTTCT	TAGTTTGTATAA	TAACATAAGTTCA	4320
Db	4261	TGACTG	CCCTTATGAT	GCCTGTCTTCT	TAGTTTGTATAA	TAACATAAGTTCA	4320
QY	4321	TTGGTT	GCATGGCATT	AGATACAAAATGGAAT	CCAAAATGTTCT	TGCATATTGATGGC	4380
Db	4321	TTGGTT	GCATGGCATT	AGATACAAAATGGAAT	CCAAAATGTTCT	TGCATATTGATGGC	4380
QY	4381	TGATC	CTTTTGATCT	CGCAGGATTTTGT	TAAAGGCAACCCGTTT	CTCGAAGGGA	4440
Db	4381	TGATC	CTTTTGATCT	CGCAGGATTTTGT	TAAAGGCAACCCGTTT	CTCGAAGGGA	4440
QY	4441	ACCTAG	TGAGATAA	TTGCTTAACAT	TGAGGCTGACGAACT	CAATGGGTTTTA	4500
Db	4441	ACCTAG	TGAGATAA	TTGCTTAACAT	TGAGGCTGACGAACT	CAATGGGTTTTA	4500
QY	4501	TACACG	AAACTTCAAGGTAA	CGAATTCCTAG	CATATTACACTTAT	TCACAGAGATTATGCA	4560
Db	4501	TACACG	AAACTTCAAGGTAA	CGAATTCCTAG	CATATTACACTTAT	TCACAGAGATTATGCA	4560
QY	4561	TTATTTT	TAAACTCTCAACT	GTTTAAACGCAT	GTGTAGATAGATTGATA	AGATTGACAAAGG	4620
Db	4561	TTATTTT	TAAACTCTCAACT	GTTTAAACGCAT	GTGTAGATAGATTGATA	AGATTGACAAAGG	4620
QY	4621	AAACTT	AGTTTATATCT	CTGGCGTTCAAAAAC	GAAAGTCCCTAGTGT	GAAATTTATCATTTT	4680
Db	4621	AAACTT	AGTTTATATCT	CTGGCGTTCAAAAAC	GAAAGTCCCTAGTGT	GAAATTTATCATTTT	4680
QY	4681	AATGTT	TAGCAGAGTACA	ATTGTTATGATTTGTT	TACGTCTATGTGCTCA	ACAGACAAGG	4740
Db	4681	AATGTT	TAGCAGAGTACA	ATTGTTATGATTTGTT	TACGTCTATGTGCTCA	ACAGACAAGG	4740
QY	4741	CTCGAG	GGATTATCTTCGAT	CAAGGCCGACAGTTAG	CTGTTGTGATAGAGGTA	ATTATT	4800
Db	4741	CTCGAG	GGATTATCTTCGAT	CAAGGCCGACAGTTAG	CTGTTGTGATAGAGGTA	ATTATT	4800
QY	4801	GCTTGT	TGTGATTGTA	ATAAGTTT	TGCTTTCAGTTT	TAAGGGTATCTAGCAA	4860
Db	4801	GCTTGT	TGTGATTGTA	ATAAGTTT	TGCTTTCAGTTT	TAAGGGTATCTAGCAA	4860
QY	4861	AAATTA	ACCTACATGCAG	ATTTACGAGGTGGCAC	CCATCGCTTTT	CATGGTAGACGTAAGA	4920
Db	4861	AAATTA	ACCTACATGCAG	ATTTACGAGGTGGCAC	CCATCGCTTTT	CATGGTAGACGTAAGA	4920
QY	4921	AAGGCT	GCTGGTGA	AACTCTTGAATAT	CACAAGGTTTATA	AAATATATATATCA	4980
Db	4921	AAGGCT	GCTGGTGA	AACTCTTGAATAT	CACAAGGTTTATA	AAATATATATATCA	4980
QY	4981	AGTTGC	ATCATTACTGT	TGTTGCGGATTAG	AGTGATATTTT	TGTTTTCGAGTT	5040

Db 4981 AGTTGCATCATTTACTGTGTTGGGATTTAGAGTGATATTTTGTGTTGGTATCGCAGTTC 5040  
QY 5041 TACAAGAAGCTATGTTGCAAACTGGAAACATAATATGAGGGCAACAGAAAGGAATACCA 5100  
Db 5041 TACAAGAAGCTATGTTGCAAACTGGAAACATAATATGAGGGCAACAGAAAGGAATACCA 5100  
QY 5101 AAGTCAGAGATTTCTCAGAACTCAATCACGTTTTCATCCCAACTTAA 5144  
Db 5101 AAGTCAGAGATTTCTCAGAACTCAATCACGTTTTCATCCCAACTTAA 5144

RESULT 2  
ABZ14979  
ID ABZ14979 standard; DNA; 2000 BP.  
XX AC ABZ14979;  
XX AC  
DT 21-JAN-2003 (first entry)  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 2784.  
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
KW Arabidopsis thaliana;  
XX Arabidopsis thaliana.  
OS  
XX WO200216655-A2.  
PN PD 28-FEB-2002.  
XX 24-AUG-2001; 2001WO-US26685.  
PF 24-AUG-2000; 2000US-227866P.  
PR 26-JAN-2001; 2001US-264647P.  
XX 22-JUN-2001; 2001US-300111P.  
PA (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.

Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

Claim 144; SEQ ID NO 2784; 577pp + Sequence Listing; English.

The invention relates to identifying a stress condition to which a plant cell has been exposed, comprising:  
(a) contacting nucleic acid representative of expressed polynucleotides in the plant cell with an array or probes representative of the plant cell genome; and  
(b) detecting a profile of expressed polynucleotides in the plant cell characteristic of a stress response. The method is useful in the production of transgenic plants, cells and seeds and in producing plants with increased tolerance to abiotic stress. The present sequence is that of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used in methods of the invention.

Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

Sequence 2000 BP; 675 A; 270 C; 353 G; 702 T; 0 other;  
Query Match 8.1%; Score 414.6; DB 24; Length 2000;  
Best Local Similarity 87.8%; Pred. No. 1e-60;  
Matches 489; Conservative 0; Mismatches 59; Indels 9; Gaps 3;  
QY 1 TTTTTCGCGGAAAATCTCGGGTTTACGTTTTCGCGGAAAATCTCGGGTTTACGTTT 60  
Db 510 TTTTTCGCGGAAAATCTCGGGTTTACGTTTTCGCGGAAAATCTCGGGTTTACGTTT 569

QY 61 TGGCGGGAAAATCTCGGGTTTACGTTTTCGCGGAAAATCTCGGGTTTACGTTTTCGTTT 120  
Db 570 TGGCGGGAAAATCTCGGGTTTACGTTTTCGCGGAAAATCTCGGGTTTACGTTTTCGTTT 628  
QY 121 CGGGAAAATCACGGATTATGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 180  
Db 629 CGGGAAAATCACGGATTATGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 688  
QY 181 TTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 240  
Db 689 TTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 748  
QY 241 TTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 300  
Db 749 TTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTTTCGTTT 808  
QY 301 TACAATAATTTGATGGTTAAATTTGGATAAACCATGGGAAACCATTAACCATTAACCATTAAC 360  
Db 809 TACAATAATTTGATGGTTAAATTTGGATAAACCATGGGAAACCATTAACCATTAACCATTAAC 868  
QY 361 TCATTTTACTCATCAAAACCAATTTGACTCATCAAAACCAATTTGACTCATCAAAACCAATTTGAG 420  
Db 869 TCATTTTACTCATCAAAACCAATTTGACTCATCAAAACCAATTTGACTCATCAAAACCAATTTGAG 928  
QY 421 TCAAAAATTTCAACTCATTTAGGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTT 480  
Db 929 TCAAAAATTTCAACTCATTTAGGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTTTCATGGTT 981  
QY 481 TTGACCCATTTTGACACCCCTTACATATGATCATTAAGTTTAAATAATCAAAAATTTACTATTG 539  
Db 982 TTGACCCATTTTGACACCCCTTACATATGATCATTAAGTTTAAATAATCAAAAATTTACTATTG 1041  
QY 540 ATAACTTTTAAACGGAA 556  
Db 1042 TAGTTTATTTTATTA 1058

RESULT 3  
ABZ17522/c  
ID ABZ17522 standard; DNA; 2000 BP.  
XX AC ABZ17522;  
XX AC  
DT 21-JAN-2003 (first entry)  
XX AC  
DE Arabidopsis thaliana stress regulated gene SEQ ID NO 5327.  
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
XX Arabidopsis thaliana.  
XX WO200216655-A2.  
XX PD 28-FEB-2002.  
XX PF 24-AUG-2001; 2001WO-US26685.  
XX PR 24-AUG-2000; 2000US-227866P.  
XX PR 26-JAN-2001; 2001US-264647P.  
XX PR 22-JUN-2001; 2001US-300111P.  
XX (SCRI ) SCRIPPS RES INST.  
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Harper JF, Kreps J, Wang X, Zhu T;  
XX WPI; 2002-304127/34.  
PT Identifying a stress condition to which a plant cell has been exposed and producing plants with increased tolerance to these abiotic stresses

PS Claim 144; SEQ ID NO 5327; 577bp + Sequence Listing; English.

XX The invention relates to identifying a stress condition to which a plant

CC cell has been exposed, comprising:

CC (a) contacting nucleic acid representative of expressed polynucleotides

CC in the plant cell with an array or probes representative of the plant

CC cell genome; and

CC (b) detecting a profile of expressed polynucleotides in the plant cell

CC characteristic of a stress response. The method is useful in the

CC production of transgenic plants, cells and seeds and in producing plants

CC with increased tolerance to abiotic stress. The present sequence is that

CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used

CC in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed

CC specification but is based on sequence information supplied to Derwent by

CC the European Patent Office.

XX Sequence 2000 BP; 681 A; 386 C; 275 G; 658 T; 0 other;

SQ

Query Match 7.8%; Score 401.4; DB 24; Length 2000;

Best Local Similarity 89.4%; Pred. No. 1.7e-58;

Matches 455; Conservative 0; Mismatches 51; Indels 3; Gaps 2;

QY 2 TTTTGGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 61

Db 1187 TTTTGGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 1128

QY 62 GCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 121

Db 1127 GCGTGAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 1069

QY 122 GCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 181

Db 1068 GCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 1011

QY 182 TCATGCTTGTATATTAAGAAATTTGGAAATATTAATTTTAAATTTAGAT 241

Db 1010 TCATGCTTGTATATTAAGAAATTTGGAAATATTAATTTTAAATTTAGAT 951

QY 242 GTGTTGGTTAAACCTAAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 301

Db 950 GTGTTGGTTAAACCTAAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 891

QY 302 ACAAAATTTGATGGTTAAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 361

Db 890 ACAAAATTTGATGGTTAAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 831

QY 362 CATTTACTCATCAACCAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 421

Db 830 CATTTACTCATCAACCAATTCGATTTAGATTTTGGTTTGGTTTATTTCAATTT 771

QY 422 CAAAAATTTCAACTCATTTAGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT 481

Db 770 CAAAAATTTCAACTCATTTAGGTTTGGTTTGGTTTGGTTTGGTTTGGTTTGGTTT 711

QY 482 TGACCCATTTTGACACCCCTACATATGAT 510

Db 710 TAACCATTTTGACATCTCTACCTGCGTT 682

RESULT 4

ABZ16852

ID ABZ16852 standard; DNA; 2000 BP.

XX

AC ABZ16852;

XX

DT 21-JAN-2003 (first entry)

XX

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4657.

XX

KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX

OS Arabidopsis thaliana.

XX WO200216655-A2.

PN

XX 28-FEB-2002.

PD

XX 24-AUG-2001; 2001WO-US26685.

PF

XX 24-AUG-2000; 2000US-227866P.

PR

XX 26-JAN-2001; 2001US-264647P.

PR

XX 22-JUN-2001; 2001US-300111P.

PR

PA (SCRI) SCRIPPS RES INST.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PA

XX Harper JF, Kreps J, Wang X, Zhu T;

PI

XX WPI; 2002-304127/34.

DR

XX Identifying a stress condition to which a plant cell has been exposed

PT

PT and producing plants with increased tolerance to these abiotic stresses

PT

XX Claim 144; SEQ ID NO 4657; 577bp + Sequence Listing; English.

PS

XX The invention relates to identifying a stress condition to which a plant

CC cell has been exposed, comprising:

CC (a) contacting nucleic acid representative of expressed polynucleotides

CC in the plant cell with an array or probes representative of the plant

CC cell genome; and

CC (b) detecting a profile of expressed polynucleotides in the plant cell

CC characteristic of a stress response. The method is useful in the

CC production of transgenic plants, cells and seeds and in producing plants

CC with increased tolerance to abiotic stress. The present sequence is that

CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used

CC in methods of the invention.

CC Note: The sequence data for this patent is not represented in the printed

CC specification but is based on sequence information supplied to Derwent by

CC the European Patent Office.

XX

SQ Sequence 2000 BP; 708 A; 255 C; 280 G; 757 T; 0 other;

Query Match 7.2%; Score 371.2; DB 24; Length 2000;

Best Local Similarity 95.3%; Pred. No. 2e-53;

Matches 404; Conservative 0; Mismatches 18; Indels 2; Gaps 2;

QY 47 CGTGTTCACGTTTTCGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTACGGG 106

Db 748 CGGTTTACGTTTTCGCGGAAATCTCGGTTTACGTTTTCGCGGAAATCTACGGA 807

QY 107 TTTACTTTTTCGCGGAAATCTACGTTTACGTTTTCGCGGAAATCTACGAGTT 166

Db 808 TTTAC-GTTTTTCGCGGAAATCTACGTTTACGTTTTCGCGGAAATCTACGGTT 865

QY 167 TACTTTTCTCAATTTTCATGTTTAAACCTTAAATTTAGAAATTTGGAAATATTAATTTAT 226

Db 866 TACTTTTCTCAATTTTCATGTTTAAACCTTAAATTTAGAAATTTGGAAATATTAATTTAT 925

QY 227 TAAATTTGGTTTAGATGTTTAAACCTTAAATTTAGAAATTTGGTTTAGAGATTTAGTTG 286

Db 926 TAAATTTGGTTTAGATGTTTAAACCTTAAATTTAGAAATTTGGTTTAGAGATTTAGTTG 985

QY 287 GTTTTATTCAATTTTACAAATTTTCATGTTTAAATTTGGTTTAGAGATTTAGTTG 346

Db 986 GTTTTATTCAATTTTACAAATTTTCATGTTTAAATTTGGTTTAGAGATTTAGTTG 1045

QY 347 CATTACAACCTCACTCAATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCA 406

Db 1046 CATTACAACCTCACTCAATTTTACTCATCAACCAATTTGACTCATCACTCACTCACTCA 1105

QY 407 TCAACTCATTTGAGTCAAAATTTTCAACTCATTTAGGTTTCACTGAGTTGAGTTGAG 466

Db 1106 TCAACTCATTTGAGTCAAAATTTTCAACTCATTTAGGTTTCACTGAGTTGAGTTGAG 1165



```
QY      467 TTGA 470
Db      1166 TTGA 1169

RESULT 5
AAF22305
ID      AAF22305 standard; DNA; 1082138 BP.
XX
AC      AAF22305;
XX
DT      20-MAR-2001 (first entry)
XX
DE      Arabidopsis thaliana chromosome 4 centromere.
XX
KW      Centromere; michromosome; vector; ds.
XX
OS      Arabidopsis thaliana.
XX
PN      WO200055325-A2.
XX
PD      21-SEP-2000.
XX
PF      17-MAR-2000; 2000WO-US07392.
XX
PR      18-MAR-1999; 99US-0125219.
PR      01-APR-1999; 99US-0127409.
PR      18-MAY-1999; 99US-0134770.
PR      13-SEP-1999; 99US-0153584.
PR      17-SEP-1999; 99US-0154603.
XX
PA      (UYCH-) UNIV CHICAGO.
XX
PI      Preuss D, Copenhaver G, Keith K;
XX
DR      WPI; 2000-587529/55.
XX
PT      Recombinant DNA construct comprising a plant centromere, useful for
PT      producing stably inherited michrosomes which can serve as vectors for
PT      the construction of transgenic plant and animal cells -
XX
PS      Claim 68; Page 977-1388; 1449pp; English.
XX
CC      The present invention relates to a recombinant DNA construct of a plant
CC      (Arabidopsis thaliana) centromere. The constructs are useful for
CC      producing stably inherited michrosomes which can serve as vectors for
CC      the construction of transgenic plant and animal cells expressing
CC      selected proteins such as hormones, enzymes, interleukins, clotting
CC      factors, cytokines, antibodies, and growth factors.
XX
SQ      Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;

Query Match      4.5%; Score 234; DB 21; Length 1082138;
Best Local Similarity 82.8%; Pred. No. 3.7e-30;
Matches 308; Conservative 0; Mismatches 50; Indels 14; Gaps 3;

QY      218 TAATTTTATAAATTGGTTTAGAATGTTGGTTAAACCTAAATTGGCATTGGTTTAGAGA 277
Db      |||||||
Db      822201 TGATTTTATAAATTGGTTTAGAATGTTGGTTAAAT-----ATTGGTTTAGAGA 822250

QY      278 TTTTAGTTGGTTTATTCAATTTTACAAAATTTGATGGTTAATTGGATAAACCATGGAA 337
Db      |||||||
Db      822251 TTTTAGTTGGTTTAAATTCAATTTTACAAAATTTGATGGTTAATTGGTAAACCACTAAA 822310

QY      338 ACCATTAA-CCATTACAACCTAACTCAATTTACTCATCAAAACCAATTGACTCATCAACTC 396
Db      |||||||
Db      822311 ACCATTAAACCCATTACACCCAACTCAATTTACTCATCAAAACCAATTGACTCATCAACTC 822370

QY      397 ATTTGACTCATCAACTCATTTGAGTCAAAAATTTCAACTCATTAGGTTTCATGGTTGAG 456
Db      |||||||
Db      822371 ATTTGACCCATCAACTCATTTGAGTCAAAAATTTCAACCCATTAGGTTTCATGGTTGAA 822430

QY      457 TTGAGTTGAGTTGACCCCATGAATTTTGACCCATTTTGACACCCCTACATAT---GATCAT 513
```

```
Db      822431 TTGAGTTGAGTTGACCCCATGAATTTTGACCCATTTTGACACCCCTACTTCTCCCAACAT 822490
QY      514 AAGTTAATAATCAAAAATTACTATTGATAAACTTTTAAACGGAATTGTTTGTAAAGTTTCA 573
Db      |||||||
Db      822491 AATTAAAAACTCTCTACAAGTTACATATATTTTTCACAAAACATTTTGAATAATAAT 822550
QY      574 TTTACGTTATTT 585
Db      |||||||
Db      822551 TGTTATCAATTT 822562

RESULT 6
ABK91084
ID      ABK91084 standard; cDNA; 1377 BP.
XX
AC      ABK91084;
XX
DT      15-NOV-2002 (first entry)
XX
DE      cDNA encoding Arabidopsis thaliana SOS2 serine/threonine protein kinase.
XX
KW      Salt overly sensitive 7; SOS2; serine/threonine protein kinase;
KW      salt tolerance; agricultural crop; rice; corn; wheat; cotton;
KW      peanut; soybean; plant protectant; plant; transgenic; gene; ss.
XX
OS      Arabidopsis thaliana.
XX
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      26..1366
FT      /*tag= a
FT      /product= "SOS2"
XX
PN      US2002095032-A1.
XX
PD      18-JUL-2002.
XX
PF      04-APR-2001; 2001US-0824735.
XX
PR      04-APR-2000; 2000US-194649P.
XX
PA      (UYAR-) UNIV ARIZONA.
XX
PI      Zhu J, Liu J, Ishitani M, Kim C, Halfter U;
XX
DR      WPI; 2002-665801/71.
DR      P-PSDB; ABG31356.
XX
PT      New SOS2 protein kinase polypeptide and nucleic acids encoding the
PT      polypeptide, useful for increasing salt tolerance to plants, especially
PT      to agricultural crops -
XX
PS      Disclosure; Fig 2; 28pp; English.
XX
CC      The present invention relates to the isolation of an Arabidopsis
CC      thaliana gene encoding salt overly sensitive 7 (SOS2) serine/threonine
CC      protein kinase. The polynucleotide sequence encoding SOS2 can be used
CC      to increase salt tolerance in plants, particularly agricultural crops
CC      such as rice, corn, wheat, cotton, peanut and soybean. The
CC      polynucleotide sequence encoding SOS2 can also be used to generate
CC      transgenic plants, as hybridization probes for RNA, cDNA and DNA to
CC      isolate those cDNAs or genes which exhibit a high degree of similarity
CC      to the sequence of the SOS2 gene, and as PCR primers for the production
CC      of DNA which encodes an enzyme having serine/threonine protein kinase
CC      activity. The present sequence encodes Arabidopsis thaliana SOS2
CC      serine/threonine protein kinase.
XX
SQ      Sequence 1377 BP; 434 A; 221 C; 341 G; 381 T; 0 other;

Query Match      4.0%; Score 206.8; DB 24; Length 1377;
Best Local Similarity 61.3%; Pred. No. 8.4e-26;
Matches 692; Conservative 0; Mismatches 2; Indels 434; Gaps 5;
```



QY 1979 TGGATCAGATATAAAGTTTGTAAAGAAATGACAAAGAAAATGAGAAGAGTGGCGAAGTACGA 2038  
Db 1 TGGATCAGATATAAAGTTTGTAAAGAAATGACAAAGAAAATGAGAAGAGTGGCGAAGTACGA 60  
QY 2039 GGTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAGTTAAAGTTTTCGAGGAACACAGA 2098  
Db 61 GGTGGTCGCACAATAGGTGAAGGAACCTTTGCTAAGTTAAAGTTTTCGAGGAACACAGA 120  
QY 2099 CACTGGTGATAATGTAGCCATCAAAAATTATGGCTAAGAGTACAATACTTTAAGAACAGAAT 2158  
Db 121 CACTGGTGATAATGTAGCCATCAAAAATTATGGCTAAGAGTACAATACTTTAAGAACAGAAT 180  
QY 2159 GGTTCATCAGGTATGTTCTGGATTGTTTTTACATCGAAACTAAGGTTTTCGCTCAATG 2218  
Db 181 GGTTCAT----- 187  
QY 2219 GTATGATCTTTGATTTCGTTTAAAGCTCTTTTACAGATAAAAAGAGAGATATCTATAATG 2278  
Db 188 -----CAGATAAAAAGAGAGATATCTATAATG 214  
QY 2279 AAGATTGTTTCGTCACCCGAAACATAGTGAGGTTGTFATGAGGTATGTTTGTGTTTCCATG 2338  
Db 215 AAGATTGTTTCGTCACCCGAAACATAGTGAGGTTGTFATG----- 251  
QY 2339 CATCTGCGAAATTTTATCTCTGAAGTGTTTTTCATCATGTTCTTCTGTTGTTTTTTTG 2398  
Db 252 ----- 251  
QY 2399 TGATTTTCCCGATGTAGGTGTTGGCGAGTCTTCGAAAATATATATATAGTTTGGAGTTTG 2458  
Db 252 -----AGGTGTTGGCGAGTCTTCGAAAATATATATATAGTTTGGAGTTTG 296  
QY 2459 TGACAGGAGGAGAGCTCTTTTGATAGAAATTTACGGAACCTCCATACTTGTAGGCAGCGTC 2518  
Db 297 TGACAGGAGGAGAGCTCTTTTGATAGAAAT----- 325  
QY 2519 CATTAGTTAAACCTCTCTACTTAATTTTAAATATATGAAATCTTTCATGCAGGTTTCAT 2578  
Db 326 -----GTTTCAT 331  
QY 2579 AAAGGGAGGCTTGAAGAAAGTGAGTCTCGAAAATACTTTCAACAGCTTGTAGATGCTGTT 2638  
Db 332 AAAGGGAGGCTTGAAGAAAGTGAGTCTCGAAAATACTTTCAACAGCTTGTAGATGCTGTT 391  
QY 2639 GCTCATTGTCACTGCAAGGGTGTTTACCACCGTGACCTAAAGGTAAGACGTTGTTTTGT 2698  
Db 392 GCTCATTGTCACTGCAAGGGTGTTTACCACCGTGACCTAA----- 431  
QY 2699 TTACCAATATTCCTCAGAATATCTCACTGCGTTGCAATCCAGACTTGATATTTTGTGTC 2758  
Db 432 ----- 431  
QY 2759 GCTATGTTATGTTATCTAGCCAGAAAATCTTTTACTCGATACAAAATGGAATCTGAAGGT 2818  
Db 432 -----AGCCAGAAAATCTTTTACTCGATACAAAATGGAATCTGAAGGT 474  
QY 2819 TTCCGATTTCCGACTCAGTGCATTGCGCTCAGGAAGTAAGTGCTTCTTATCTCTGCTTCAGC 2878  
Db 475 TTCGGATTTCCGACTCAGTGCATTGCGCTCAGGAA----- 508  
QY 2879 AGTCTGCTTACGTGGTCATTAACTTGTTATATATACTCAATCAGGAGTAGAACCTTCTGCGT 2938  
Db 509 -----GGAGTAGAACTTCTGCGT 526  
QY 2939 ACCACATGTGGAACCTCCGAACATATGTAGCTCCAGAGGTACTTAGTGACAGGGTTACGAT 2998  
Db 527 ACCACATGTGGAACCTCCGAACATATGTAGCTCCAGAGGTACTTAGTGACAGGGTTACGAT 586  
QY 2999 GGTTCAGCAGCTGATATTTGGTCTTGGGGGTTATTTCTTTTCGTTATATTGGCTGGATAT 3058  
Db 587 GGTTCAGCAGCTGATATTTGGTCTTGGGGGTTATTTCTTTTCGTTATATTGGCTGGATAT 646  
QY 3059 TTACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAAAAGTAAGT 3106

Db 647 TTACCTTTTCCGAGACGGATCTTCCAGGTTGTACAGAAAATAAAT 694

RESULT 7  
AAC50382  
ID AAC50382 standard; DNA; 1341 BP.  
XX AAC50382;  
AC AAC50382;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 64633.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; terminator; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP10333405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.



QY 3039 TCGTTATATGGCTGGATATTACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAA 3098  
Db 602 TCGTTATATGGCTGGATATTACCTTTTCCGAGACGGATCTTCCAGGGTTGTACAGAA 661  
QY 3099 AAGTAAGT 3106  
Db 662 AAATAAAT 669  
RESULT 8  
AAF22306  
ID AAF22306 standard; DNA; 163319 BP.  
XX  
AC AAF22306;  
XX  
DT 20-MAR-2001 (first entry)  
XX  
DE Arabidopsis thaliana chromosome 4 centromere.  
XX  
KW Centromere; michrosome; vector; ds.  
XX  
OS Arabidopsis thaliana.  
XX  
PN WO200055325-A2.  
XX  
PD 21-SEP-2000.  
XX  
PF 17-MAR-2000; 2000WO-US07392.  
XX  
PR 18-MAR-1999; 99US-0125219.  
PR 01-APR-1999; 99US-0127409.  
PR 18-MAY-1999; 99US-0134770.  
PR 13-SEP-1999; 99US-0153584.  
PR 17-SEP-1999; 99US-0154603.  
XX  
PA (UYCH-) UNIV CHICAGO.  
XX  
PI Preuss D, Copenhaver G, Keith K;  
XX  
DR WPI; 2000-587529/55.  
XX  
PT Recombinant DNA construct comprising a plant centromere, useful for  
PT producing stably inherited michrosomes which can serve as vectors for  
PT the construction of transgenic plant and animal cells -  
PS Claim 79; Page 1389-1451; 1449pp; English.  
XX  
CC The present invention relates to a recombinant DNA construct of a plant  
CC (Arabidopsis thaliana) centromere. The constructs are useful for  
CC producing stably inherited michrosomes which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
SQ Sequence 163319 BP; 53475 A; 30117 C; 28735 G; 50992 T; 0 other;  
Query Match 3.2%; Score 163.6; DB 21; Length 163319;  
Best Local Similarity 53.7%; Pred. No. 2.2e-18;  
Matches 404; Conservative 0; Mismatches 319; Indels 29; Gaps 2;  
QY 165 TTTACTTTTCTCAATTCATGCTTGGTTAAACCTTAATTGGCATTGGTTAGAGATTTTAGT 224  
Db 120103 TTTAGTTGATAATAAATATACGGATGGTGGATGAAGCCATTGTGAAAAAAGATTTTGGAT 120162  
QY 225 ATTAAATGGTTTAGATGCTTGGTTAAACCTTAATTGGCATTGGTTAGAGATTTTAGT 284  
Db 120163 ATTCAATCGGTTTAGAAAT-TTGTGAAACTCAAAATGATATCGGTTTAGAGATTTAAGT 120221  
QY 285 TGGTTTATTCAATTTTACAAAATTTTGATGGTTAATTGGATAAACCATGGAACCATTA 344  
Db 120222 TTGTTAATTCATTTTACAAAATTTGGTGGTT----- 120255  
QY 345 ACCATTACAACCTAACTCATTTTACTCATCAACCAATTGACTCATCACTCATTTGACT 404

Recombinant DNA construct comprising a plant centromere, useful for producing stably inherited michrosomes which can serve as vectors for the construction of transgenic plant and animal cells -

Claim 79; Page 1389-1451; 1449pp; English.

The present invention relates to a recombinant DNA construct of a plant (Arabidopsis thaliana) centromere. The constructs are useful for producing stably inherited michrosomes which can serve as vectors for the construction of transgenic plant and animal cells expressing selected proteins such as hormones, enzymes, interleukins, clotting factors, cytokines, antibodies, and growth factors.

Sequence 163319 BP; 53475 A; 30117 C; 28735 G; 50992 T; 0 other;

Query Match 3.2%; Score 163.6; DB 21; Length 163319;  
Best Local Similarity 53.7%; Pred. No. 2.2e-18;  
Matches 404; Conservative 0; Mismatches 319; Indels 29; Gaps 2;  
QY 165 TTTACTTTTCTCAATTCATGCTTGGTTAAACCTTAATTGGCATTGGTTAGAGATTTTAGT 224  
Db 120103 TTTAGTTGATAATAAATATACGGATGGTGGATGAAGCCATTGTGAAAAAAGATTTTGGAT 120162  
QY 225 ATTAAATGGTTTAGATGCTTGGTTAAACCTTAATTGGCATTGGTTAGAGATTTTAGT 284  
Db 120163 ATTCAATCGGTTTAGAAAT-TTGTGAAACTCAAAATGATATCGGTTTAGAGATTTAAGT 120221  
QY 285 TGGTTTATTCAATTTTACAAAATTTTGATGGTTAATTGGATAAACCATGGAACCATTA 344  
Db 120222 TTGTTAATTCATTTTACAAAATTTGGTGGTT----- 120255  
QY 345 ACCATTACAACCTAACTCATTTTACTCATCAACCAATTGACTCATCACTCATTTGACT 404

RESULT 9

ABZ17294/c

ID ABZ17294 standard; DNA; 2000 BP.

XX

AC ABZ17294;

DT 21-JAN-2003 (first entry)

XX

DE Arabidopsis thaliana stress regulated gene SEQ ID NO 5099.

XX

KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX

OS Arabidopsis thaliana.

XX

PN WO200216655-A2.

XX

PD 28-FEB-2002.

XX

PF 24-AUG-2001; 2001WO-US26685.

XX

PR 24-AUG-2000; 2000US-227866P.

PR 26-JAN-2001; 2001US-264647P.

PR 22-JUN-2001; 2001US-300111P.

XX

PA (SCRI ) SCRIPPS RES INST.

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX

PI Harper JF, Kreps J, Wang X, Zhu T;

XX

DR WPI; 2002-304127/34.

XX

PT Identifying a stress condition to which a plant cell has been exposed

PT and producing plants with increased tolerance to these abiotic stresses

Db 120256 --CATTGCACCCAACTTATTTGATCAATATAACCACTTGGACTCAATACTTATTTGAAC 120313  
QY 405 CATCAACTCATTTGAGTCAAAAATTCAACTCATTTAGGGTTCATGGGTTGAGTTGAGTTG 464  
Db 120314 CATCAACTCATTTGAGTCAAAAATTCAACTCATTTAGGGTTCATGGGTTGAGTTGAGTTG 120373  
QY 465 AGTTGACCCATGAATTTTGACCCCATTTTGACACCCCTACATATGATCATATAAGTTAAT 524  
Db 120374 GGTGATCCATGAAACTTAAACCATTTTGACAAACCCCTATAAAAAACAAT 120433  
QY 525 CAAAAATTACTATTGATAAATTTTAAACGGAATTTGTTTGAAGTTTCATTTACGTTATT 584  
Db 120434 TGAAAAATAGAGTAAAGTGTATTTTCCCAATGTGACCTTTTATTTATAGATACTTATTA 120493  
QY 585 TATATAAAAAACATAAATGCAAAAGTAAATGTATAGTTAATTTTATTTTAAATAATGC 644  
Db 120494 TGATAAACCACTGCACTATATGCAAGACAAACAGTTAATAAATAATTTTAATGTATTG 120553  
QY 645 AAAAAATTTACTGTAATACTTTTTCATTTCTTATCAATATTTTCTTATATTTCAATTAGC 704  
Db 120554 ACGATAATGAAGGTGTTCTATTTTATATTTTACATAAACGATAGTGTATTAATGATTA 120613  
QY 705 CACCAACACCTACATTTTCATATTTCTCTTTCACATTTCAATTTTCTTCTTCTTCTTCC 764  
Db 120614 TCAAGATTAGGTAAATTTTATTTTATTTTATTTTCTAATGTTTATGTTTATTTATGCTCTCT 120673  
QY 765 TTCTTATCATCTTCATTTGCCAATTTTTCATTTGACATTTGTCACTTCTTCTTCTTCTTCTTCT 824  
Db 120674 ATATATATATATATATATATCTTAAATAGATACACAAATGTTAAGTATTTTCTTTTGTAT 120733  
QY 825 ATTTTGAACATTTAATGTGAATCCCAACCGACTATAGGACTATCAAGAAAGTTTTCAAAC 884  
Db 120734 AGATTTGTATATTTTAAATGCAATATTAATCCTAGTTTGTGTTTAAATAATTAATGAGTAA 120793  
QY 885 TTTTAAAAAAGATCTTTGAACCTTTTAAAGCAA 916  
Db 120794 ATTTAAAGCAGAAATTAATTGGAGAAAAAAA 120825





CC producing stably inherited mitochondria which can serve as vectors for  
CC the construction of transgenic plant and animal cells expressing  
CC selected proteins such as hormones, enzymes, interleukins, clotting  
CC factors, cytokines, antibodies, and growth factors.  
XX  
SQ Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;  
Query Match 2.2%; Score 112.4; DB 21; Length 1082138;  
Best Local Similarity 81.2%; Pred. No. 1e-09;  
Matches 143; Conservative 0; Mismatches 31; Indels 2; Gaps 1;  
QY 1 TTTTGGCGGAAATCTCGGGTTACGTTTTCGCGGAAATCTCGGTTTACGTTT 60  
Db 678743 TTTATGGTAGAAATACGGGTTTATGTTTGGCAAGCTAATTAGGGTTTACGTTT 678684  
QY 61 TGGCGGAAATCTCGGGTTACGTTTTCGCGGAAATCTCGGTTTACTTTT 118  
Db 678683 TGGCGGAAATCTCGGGTTACGTTTTCGCGGAAATCTCGGTTTACTTTT 678624  
QY 119 TGGCGGAAATCTCGGGTTACGTTTTCGCGGAAATCTCGGTTTACTTTT 174  
Db 678623 TGGCGGAAATCTCGGGTTACGTTTTCGCGGAAATCTCGGTTTACTTTT 678568  
RESULT 12  
ID ABZ10246  
XX  
AC ABZ10246; standard; DNA; 8056 BP.  
XX  
DT ABZ10246;  
XX  
TX 16-JAN-2003 (first entry)  
DE Haematopoietic cell proliferation disorder related DNA sequence #386.  
XX  
KW Human; haematopoietic cell proliferation disorder; cytostatic;  
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
KW cytosine methylation state; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO20027272-A2.  
XX  
XX 03-OCT-2002.  
PF 26-MAR-2002; 2002WO-EP03401.  
XX  
XX 26-MAR-2001; 2001US-278333P.  
PR  
XX (EPIG-) EPIGENOMICS AG.  
PA  
XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;  
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
PI Pelet C, Schwöpe I, Ziebarth H;  
XX  
DR WPI; 2003-018942/01.  
XX  
XX Detecting and differentiating between hematopoietic cell proliferative  
PT disorders, comprises contacting a target nucleic acid with a reagent  
PT that distinguishes between methylated and non-methylated CpG  
PT dinucleotides -  
XX  
XX Claim 28; SEQ ID 386; 117pp; English.  
PS  
XX  
CC The present invention describes a method for detecting and  
CC differentiating between haematopoietic cell proliferative disorders  
CC associated with at least 1 gene and/or their regulatory regions in a  
CC subject. The method comprises contacting a target nucleic acid in a  
CC biological sample obtained from the subject with at least 1 reagent,  
CC which distinguishes between methylated and non-methylated CpG  
CC dinucleotides within the target nucleic acid. AB209861 to AB211118  
CC represent specifically claimed nucleotide sequences from the present  
CC invention. Oligonucleotides from the present invention can be used: for

CC differentiating between healthy haematopoietic cells and proliferative  
CC disorder haematopoietic cells; for differentiating between acute  
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
CC determining the cytosine methylation state and/or single nucleotide  
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
CC related sequences and their complements; and as primers for the  
CC amplification of haematopoietic cell proliferation disorder related  
CC DNA sequences. The nucleotide sequences from the present invention can  
CC also be used for detecting a predisposition to, differentiation between  
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
CC haematopoietic cell proliferation disorders. The present method enables  
CC a highly specific classification of haematopoietic cell proliferative  
CC disorders allowing for improved and informed treatment of patients.  
XX  
SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;  
Query Match 2.0%; Score 105; DB 25; Length 8056;  
Best Local Similarity 45.5%; Pred. No. 1.3e-08;  
Matches 584; Conservative 0; Mismatches 675; Indels 24; Gaps 5;  
QY 165 TTTACTTTTCTCAATTTTCATTTGTTGTTAAACCTAAATTTGGATTGGTTTAGAGATT 224  
Db 1091 TTTAATGTTTAAAAATGTTATATATTTGTTTAAATTTTAAATTTTAAATTTTAAATTT 1150  
QY 225 ATTAATTTGGTTTAGATGTTGTTAAACCTAAATTTGGATTGGTTTAGAGATT 284  
Db 1151 TAAAAATTAATATATTTTGGATTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1210  
QY 285 TGGTTTATTTCAATTTTACAAAATTTGATGGTTAAATTTGGATAAACCATGGAACCATTA 344  
Db 1211 TAAATTTTAAATAAATAAATAATTTTGAATAAAAAAATGTTATGTAATTAAT 1270  
QY 345 ACCATTACAACTTAATCTCAATTTTACTCATCAAAACCAATTTGACTCATCACTGACT 404  
Db 1271 AATATATAAAATGTAATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1330  
QY 405 CATCAACTCATTTGAGTCAAAAATTTCAACTCATTTAGGGTTTCATGGTTGAGTTGAGTTG 464  
Db 1331 TTTTAAATAAATAAATAAATAATTTTAAATTTTAAATAAATAAATAAATAAATAAATAA 1389  
QY 465 AGTTGACCCCATGAATTTTGACCCCATTTTGACACCCCTACATATGATCATAGTTAATAAT 524  
Db 1390 AAAATAATATTTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1449  
QY 525 CAAAAATTTACTATTGATAAATTTTAAACCGGAATTTGTTTGTAGTTTCACTTTACGTTAT 584  
Db 1450 ATTATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1509  
QY 585 TATATAAAAAACATAATGCAAAAGTACTAATGTATAGTTATTTTATTTTAAATAATGC 644  
Db 1510 ATAAATAATAATAATGTTAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAA 1569  
QY 645 AAAAATAATTACTGTAAATACTTTTTCATTTCTTATCAATATTTTCTTATATTCATAGC 704  
Db 1570 TTAAAT-TTAAATTTTAAATTTTATATTTTAAATTTTAAATAAATAAATAAATAAATAA 1628  
QY 705 CACCAACACCTACATTTTCAATTTCTCTTTCATTTTCAATTTTCTTCTTCTTCTTCTT 764  
Db 1629 TGAATTTGAAAATTTAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1688  
QY 765 TTCTTATCATCTTTCATTTGCCAATTTTTCATTTGACATTTGATCGTTTACTTTTGTATG 824  
Db 1689 AATTATAAAAAATTAATTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1748  
QY 825 ATTTTGAACATTTAATGTGAATCCCAACCGACTATAGGACTATCAAGAAGTTTTCAAAC 884  
Db 1749 TAAATAAAAAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1808  
QY 885 TTTTAAAAAAGATCTTGAACCTTTAAAGCAAAATCCCAACCGACTATAGAACCATAGTCC 944  
Db 1809 TTTAATAAATAAATAAATAAATTTTGTAAAAAATAAATAAATAAATAAATAAATAAATAA 1868  
QY 945 TCTTCTTGTGTGAAGCTCTTCTCTGCGCGGAAATCTAGGCCATAAAAGCCTCTT 1004

Db 1869 AATTAATTAATTTTATTTTATTAATAAAAAATTTTATATAAAAAATTTTATATAAAAAATTAAT 1928

QY 1005 CAACATCACCTAGTATATTGACCGTGACCACTCTTTTGTGACCACTGCTTTGTGAATGAACC 1064

Db 1929 TAAATTATAATAAATAAAAAATTAATAATTTTAAATAAAAAATTAATAATTAATTT 1988

QY 1065 GTCGATAAACCGTGTATCACTTACGCCAAATTTTCCCTAGTGTGTTTCCAAACTCT 1124

Db 1989 TTATTTAAATTAATAAAAAATTTTAAATATATTTTAAATAAAAAATTAATAATTTT 2048

QY 1125 CACGAATCCTTACGAACCTTTTATATATCACTTTGTAGCACTTGGAAGTATCTTTGTA 1184

Db 2049 ----TTTATTTTAAATAAAAAATTTTATTAATAAAAAATTTTAAATAAAAAATTAATAATTA 2104

QY 1185 TGCTTTGCTTAACTTAGACATCCTTGTCTCTTGTGTTTGTGTTTGTGTTGTTGTTGTTGTT 1244

Db 2105 AATTATTTAATAAATTTTAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2164

QY 1245 AA-----ATGAAGTTAAATAATTTGTAGTTAAATAATTAATAATTTTACTA 1288

Db 2165 AAAAAATTATAATTTTATTAATAATTTTAAATAAAAAATTAATAAAAAATTAATAAAAAATTA 2224

QY 1289 ATTGAGTTGCTTAAATCATAGTCTAGATAATTTGAAAAAATTAATAAAAAATTTTGA 1348

Db 2225 ATTAAATTTTAAATAAAAAATTTATTTTAAATAAAAAATTAATAAAAAATTTTAAATTT 2284

QY 1349 ATACTATATGTTTATTTTAAATAATTTTAAATAATTTGATAAGTGATATTAGATTAGT 1408

Db 2285 TTTTAAATTTTATTTTATTTT--AAATAATTTTATTTTATTTTGTATTTAAATAATTTTAT 2342

QY 1409 TTTTCTTTTCTTTTAAATTTT 1431

Db 2343 TTAATTTTAAATAATAATTTT 2365

RESULT 13

ABX21874

ID ABX21874 standard; cDNA; 404 BP.

XX ABX21874;

DT 10-FEB-2003 (first entry)

DE Human GDP-mannose 4,6-dehydratase (GM4,6D) DNA #3931.

XX Human; GDP-mannose 4,6-dehydratase; GM4,6D; gene; ss; inflammation; cellular fucosylation; glycoconjugate fucosylation; transplant rejection; arthritis; asthma; sepsis; reperfusion injury; stroke; infection; complex carbohydrate; gene replacement therapy; immunosuppressive; antiinflammatory; antiarthritic; antibacterial; cerebroprotective; antiasthmatic; vasotropic.

XX Homo sapiens.

OS US2002110548-A1.

XX 15-AUG-2002.

XX 11-JUN-2001; 2001US-0878574.

XX 22-NOV-1996; 96US-0753233.

PR 03-DEC-1997; 97US-0984246.

PR 09-SEP-1998; 98US-0149674.

PR 14-JUN-1999; 99US-0333177.

XX (GEMY ) GENETICS INST INC.

PA Sullivan F, Kriz R, Kumar R;

XX WPI; 2003-066673/06.

XX New composition comprising GDP-mannose 4,6-dehydratase (GM4,6D)

PT peptide, for manufacturing complex carbohydrates, or as targets for screening GM4,6D antagonists for treating e.g. arthritis, or transplant rejection

PT rejection

XX Disclosure; SEQ ID NO 3933; 6pp; English.

PS The invention relates to a composition comprising a human GDP-mannose 4,6-dehydratase (GM4,6D) peptide. The peptide is useful for identifying GM4,6D inhibitors. GM4,6D inhibitors are useful for reducing inflammation in a mammalian subject and for treating or ameliorating diseases affected by the level of cellular fucosylation or diseases affected by the fucosylation of glycoconjugates. These diseases include arthritis, transplant rejection, asthma, sepsis, reperfusion injury, stroke or infection. The GM4,6D peptide or a polynucleotide encoding it is also useful for manufacturing complex carbohydrates and as targets for screening small molecule antagonists of the activity of the enzyme. The polynucleotide is useful in developing an assay for defects in the enzyme, as well as in gene replacement therapy. Sequences ABX17942-ABX17944 and ABX17947-ABX33716 represent DNA molecules encoding human GM4,6D peptides of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 404 BP; 113 A; 82 C; 87 G; 122 T; 0 other;

SQ Query Match 1.9%; Score 97.6; DB 25; Length 404; Best Local Similarity 64.7%; Pred. No. 1.8e-07; Matches 145; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 2916 ATCAGGAGTAGAACTTCTCGTACCACATGTGGAACCTCCGAACATGCTAGCTCCAGAGG 2975

Db 78 AACAGGAGGATGAACCTCTCCGCACTGCTTGTGGAACCTCCAAATATGTTGCTCCTGAGG 137

QY 2976 TACTTAGTGACAGGTTACGATGGTTACGAGCTGATATTTGGTCTTCCGGGTTATTC 3035

Db 138 TGCTTAATGATAGAGTTATGTTGGTTCTACATCTGATATCTGGTCTGTGGAGTCATTC 197

QY 3036 TTTTCGTTATATTGGCTGGATATTTACCTTTTCCGAGACGGATCTCCAGGTTGTACA 3095

Db 198 TCTTTGTGCTTATGGCTGGGTACTTGGCCCTTTGATGAGCCAAATCATGCGACACTGTACC 257

QY 3096 GAAAAAGTAAGTAACATATCTTTTCGGGAAGAAATCATGAATTCCT 3139

Db 258 AAAAAATTTGGCAGGGCCCAATTCACATGTCCATCATGTTTCT 301

RESULT 14

ABX62736

ID ABX62736 standard; DNA; 397 BP.

XX ABX62736;

AC ABX62736;

XX 25-FEB-2003 (first entry)

DT Arabidopsis thaliana expressed sequence related polynucleotide #851.

DE Transgenic plant; plant; genetically modified cell; environmental stress; ribozyme creation; disease resistance; stress tolerance; fungicide screening; insecticide screening; gene; ds.

XX Arabidopsis thaliana.

OS US2002040490-A1.

XX 04-APR-2002.

XX 26-JAN-2001; 2001US-0770423.

PF 27-JAN-2000; 2000US-178512P.

XX (GORL/) GORLACH J.

PA (ANYV/) AN Y.  
PA (HAMI/) HAMILTON C M.  
PA (PRIC/) PRICE J L.  
PA (RAIN/) RAINES T M.  
PA (YUYV/) YU Y.  
PA (RAME/) RAMEAKA J G.  
PA (PAGE/) PAGE A.  
PA (MATH/) MATHEW A V.  
PA (LEDF/) LEDFORD B L.  
PA (WOES/) WOESSNER J P.  
PA (HAAS/) HAAS W D.  
PA (GARC/) GARCIA C A.  
PA (KRIC/) KRICKER M.  
PA (SLAT/) SLATER T.  
PA (DAVI/) DAVIS K R.  
PA (ALLE/) ALLEN K.  
PA (HOFF/) HOFFMAN N.  
PA (HURB/) HURBAN P.  
XX  
PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;  
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;  
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;  
PI Hurban P;  
XX WPI; 2003-110411/10.  
DR  
XX Novel Arabidopsis thaliana nucleic acid useful for identifying  
PT homologous or related genes, and to create genetically modified and  
PT transgenic organisms, such as plant cells and plants -  
XX  
PS Claim 1; SEQ ID NO 851; 43pp; English.  
XX  
CC The invention describes an Arabidopsis thaliana nucleic acid (I). The  
CC polypeptide (II) encoded by (I), transgenic plant (III) or genetically  
CC modified cell (IV) are useful for screening a candidate agent for its  
CC biological effect, by combining the candidate agent with (II), (III) or  
CC (IV), and determining the effect of the candidate agent on (II), (III) or  
CC (IV). (I) is useful for identifying homologous or related genes, for  
CC producing compositions that modulate the expression or function of its  
CC encoded protein, for mapping functional regions of the protein, in  
CC diagnosis, for studying associated physiological pathways, for genetic  
CC manipulation of cells, preferably plant cells, in screening assays of  
CC various plant strains to determine the strains that are capable of  
CC withstanding a particular disease or environmental stress, for enhancing  
CC or inhibiting production of biosynthetic product in a plant, for  
CC producing polypeptides, as probes for the detection of mRNA in biological  
CC samples, to generate additional copies of (I), to generate ribozymes or  
CC oligonucleotides, as single stranded DNA probes or as triple-strand  
CC forming oligonucleotides, and to create genetically modified and  
CC transgenic organisms, such as plant cells and plants. (II) or (III) is  
CC useful for introducing or improving disease resistance and stress  
CC tolerance in plants, screening biological active agents, e.g.,  
CC fungicides, insecticides, etc., and for elucidating biochemical pathways.  
CC (III) is useful as crops for their enhanced disease resistance, enhanced  
CC traits of interest, for screening programs, as crops which exhibit  
CC enhanced tolerance to environmental stress, or to produce a factor. This  
CC sequence represents a nucleic acid that may correspond to naturally  
CC occurring Arabidopsis thaliana expressed sequences.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC US patent office at  
CC seqdata.uspto.gov/sequence.html?DocID=999909770423.  
XX  
SQ Sequence 397 BP; 114 A; 81 C; 81 G; 121 T; 0 other;

Query Match 1.9%; Score 97.2; DB 25; Length 397;  
Best Local Similarity 69.5%; Pred. No. 2e-07;  
Matches 132; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2910 TACTCAATCAGGAGTAGAACTTCTCGGTACACATGTGGAACTCCGAACTATGATGCTC 2969  
Db 28 TACCTGAACAAGGAGTTACCATCTCTAAAGACACATGTGGAACTCCCAATTACGTTGCTC 87

QY 2970 CAGAGGTACTTAGTGGACAGGGTTACGATGGTTTCAGCAGCTGATATTTGGTCTTGGGGG 3029  
Db 88 CTGAGGTTCTCAGTCAAGGGTTACAATGGTGCCGTTGCAGATATTTGGTCTTGGGGG 147  
QY 3030 TTATTCTTTTCTGTTATATTGGCTGGATATTTACCTTTTCCGAGACGGATCTTCCAGGCT 3089  
Db 148 TCATCCTTTATGTTCTTATGGCAGGATATCTTCCATTTGATGAATGGATCTACCAACTT 207  
QY 3090 TGTACAGAAA 3099  
Db 208 TATATAGTAA 217  
RESULT 15  
AAC42994  
ID AAC42994 standard; DNA; 1338 BP.  
XX  
AC AAC42994;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37631.  
XX  
KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway;  
KW metabolic pathway; promoter; termination sequence; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 1.9%; Score 97.2; DB 21; Length 1338;  
Best Local Similarity 69.5%; Pred. No. 2.2e-07;  
Matches 132; Conservative 0; Mismatches 58; Indels 0; Gaps 0;



Search completed: October 21, 2003, 01:44:08  
Job time : 1224 secs